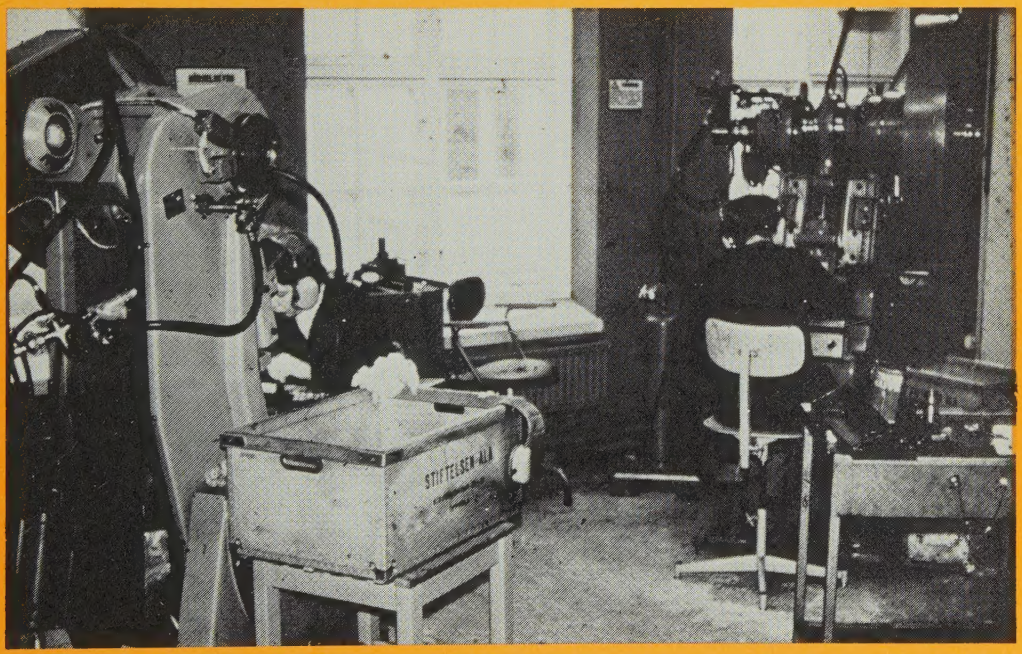
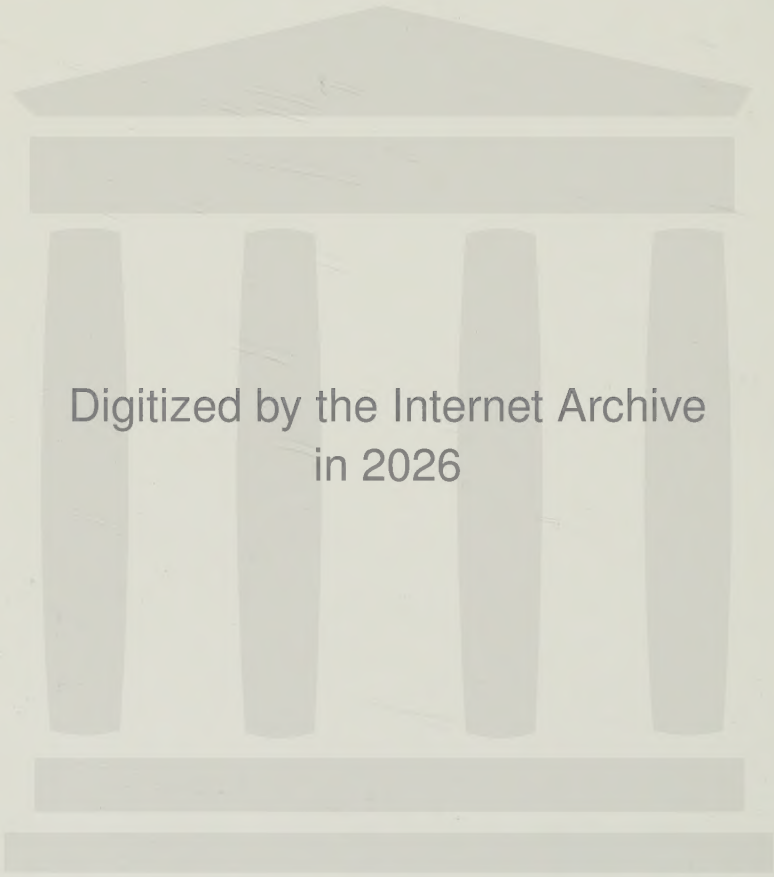
Uppsala 1974

Rikard Palmér

# Prediction of Work Performance and Work Adjustment in Mentally Retarded Adults



SCANDINAVIAN UNIVERSITY BOOKS


Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41548914_0008

SCANDINAVIAN UNIVERSITY BOOKS

DANMARK MUNKSGAARD *København*
NORGE UNIVERSITETSFORLAGET *Oslo, Bergen, Tromsø*
SVERIGE ESSELTE STUDIUM *Stockholm, Göteborg, Lund*

# Prediction of Work Performance and Work Adjustment in Mentally Retarded Adults

Inaugural Dissertation

by

Rikard Palmér

Fil.lic., Värml.

By due Permission
of the Faculty of Social Sciences
of the University of Uppsala,
to be Publicly Discussed in Lecture Hall IV,
on May 10th, 1974, at 10 a.m.,
for the degree of
Doctor of Philosophy

UPPSALA 1974

UNIV.
LEIDEN
BIBL.

Rikard Palmér

# Prediction of Work Performance and Work Adjustment in Mentally Retarded Adults

ESSELTE STUDIUM

Translated by Neil Tomkinson, B.A.,
with the assistance of Jean Gray, B.A., teol. kand.

UNIV.
LEIDEN
BIBL.

© Rikard Palmér 1974
ISBN 91-24-25148-8

Mångfaldigande av innehållet i denna bok, helt eller delvis är enligt lag om upphovsrätt av den 30 december 1960 förbjudet utan medgivande av förlaget. Förbudet gäller varje form av mångfaldigande genom tryckning, duplicering, stencilering, bandinspelning etc.

Printed in Sweden by
TEXTGRUPPEN I UPPSALA AB 1974

ISBN 91–7222–067–8

*To my Family*

# Acknowledgements

The facilities required for these investigations and, in the initial stage, economic support also were provided by the ALA Foundation, which realized that scientific data of an interdisciplinary nature are of vital importance for successful habilitation of the mentally retarded. A stimulating and fruitful co-operation developed between the researchers and practitioners closely associated with the sheltered workshop run by the Foundation from 1966 to 1970 in Uppsala. This was due to a great extent to the contributions of Carl-Eric Ohlsson, the manager of the workshop, Nils Håkansson, the chief administrator of the Foundation, Christina Hedsjö, welfare officer, Fil. lic. Gunnar Holm, research assistant, Jan-Olof Lööw, research technician, Lotti Söder, secretary, and the supervisors, Margit Carlsson, Sven Esperi, Göthe Jansson, Kjell Lindberg, Einar Meland and Arne Sjöberg.

The original planning of the research activities associated with the ALA Foundation, of which these studies form a part, was directed by Prof. Ragnar Berfenstam, the head of the Department of Social Medicine, University of Uppsala, working together with Dr. Fredrik Fränkel and Docent Lars Kebbon, all of whom also contributed in other ways to the completion of the studies. Docent Lars-Erik Jönsson served as physician and medical adviser at the workshop and was responsible for evaluating and classifying the medical information available on each subject. Chief engineer Gert Markström acted as technological expert and carried out MTM analyses of the tasks included in the investigation. Docent Gunnar Jansson and Docent Bengt Nordgren, Fil. lic. Sonja Calais, Fil. lic. Gunnar Kylén and Fil. lic. Maja Witting took part in repeated consultations and discussions.

After the activity at the sheltered workshop closed down for various reasons, I was enabled by the generosity of Prof. Karl-Georg Ahlström to continue my work at the Department of Educational Research of the Uppsala Teacher Training College, where physical facilities and the necessary assistance were placed at my disposal. Long before this, however, Prof. Ahlström had in fact become my closest mentor and guide. In this capacity he resolved many of the problems which cropped up and unreservedly shared his outstanding knowledge of research methods in the behavioural sciences.

Many other teachers, colleagues and friends have also contributed in various ways. The scientific approach in these investigations was markedly influenced by the basic training in scientific method I received from Prof. Wilhelm Sjöstrand

and the late Prof. Rudolf Anderberg. A cross-validation study reported in this thesis was carried out under the guidance of Prof. Ingmar Dureman, who in addition offered valuable suggestions and opinions on the treatment of the data collected. I was able to consult Prof. K. G. Jöreskog and Assoc. Prof. Paul Seeger concerning certain more complicated statistical problems. Fil. lic. Lars Bäckström and Bo Jacobsson, computer specialist for the Department of Educational Research, acted as consultants on computer analysis. On a number of occasions, I discussed various psychometric and general methodological problems with Prof. Sten Henrysson and Assoc. Prof. Gösta Berglund, Docent Karl Henrik Eriksson, Docent Olle Österling and Fil. lic. Bengt Ramund.

The translation into English was prepared by Neil Tomkinson, B.A., with the assistance of Jean Gray, B.A., *teol. kand.* I was also able to call on Nancy Belmore, researcher at the Department of Educational Research, for occasional advice on matters of style and terminology.

The former head of the Domestic Science College in Uppsala, Karin Holmvall, and its present director, Gunvor Fridén, generously consented to my taking long periods of leave from my permanent post there.

Research support was granted by the Bank of Sweden Tercentenary Fund, the National Board of Education and the National Mayday Flower Fund.

I wish to express my heartfelt gratitude to all these people and organisations.

First and foremost, however, my thanks are due to my wife, Margareta, and my daughters, Elisabet, Monika, Anne and Birgitta, who have given me the courage to persist when difficulties accumulated and who have tolerated my necessary neglect of our life together as a family, which is, after all, what I value most.

Uppsala, February 1974 *Rikard Palmér*

# Contents


Acknowledgements
Contents
Tables
Figures

Chapter 1. The Background 1
1. Society's care for the mentally retarded 1
2. The potential för adjustment to work and the need of individual guidance 3
3. Methodological problems in previous prediction studies 6
4. Modern lines of methodological development 8
A. The combination of prediction variables 8
B. The analysis of the problems of criteria 11
C. Prediction studies among the moderately retarded 17
D. Cross-validations and replications 18
5. The author's own empirical studies 19
A. Summary of the initial position 19
B. Aims 20
C. The concrete conditions for carrying out the studies 20
D. The terminology employed 21

Chapter II. The Recruitment and Composition of the Experimental Groups 22
1. Etiological classification as principle of selection 22
2. Inventory of needs and criteria for selection 23
3. The recruiting procedure and the problem of generalization 26
4. Sample characteristics 27

Chapter III. Prediction Variables: Selection and Results of Measurements 32
1. Psychomotor, spatial-inductive and numerical abilities 32
A. Choice of test instruments 32
B. Test administration and evaluation of performances 34
C. Estimation of reliability 36
D. Average performance level 37
E. Ability to differentiate and forms of distribution 40
F. Sex differences and intercorrelation patterns 41
2. Personality characteristics 45
A. Measuring problems 45
B. Analysis of sub-scales in a general observation schedule of the check-list type 47
C. Analysis of sub-scales in a five-degree rating schedule 51

Chapter IV. Criterion Variables: Selection and Results of Measurements 52
1. Ratings of working capacity 52
A. The prognosis-group criterion 53
B. The percentage-group criterion 54
C. The wages-group criterion 55
D. Correlations between the criterion variables 57

2. The measurement of work performances 57
A. The selection of tasks 57
B. The combination of the tasks to form the testing programme 59
C. The gradual implementation of the testing programme 62
D. The nature of the tasks and measurement of qualitative performance 65
Requirements for the approval of finished products 65
The degree of difficulty of the tasks 65
Qualitative development of performance 67
Measurement of qualitative performance 68
E. Measurement of quantitative performance in the eight tasks 69
Quantity and quality 69
Quantitative development of performance 71
Measurement of quantitative performance 75
Estimation of reliability 77
Sex differences and intercorrelations 77
Dimensions of work performance 81
Composite measurement of work performance 84
3. Observations of behaviour and adjustment during the testing programme 85
4. Summary analysis of the criterion variables 91

Chapter V. Prediction of Performance at Work and Adjustment to Work 99
1. Prediction of rated working capacity 100
A. Product-moment correlations between ability variables and rating criteria 100
B. Prediction of rating criteria from combinations of ability variables 102
How to find "the best regression equation" 102
Pitfalls in stepwise multiple-regression analyses 104
The results of the analyses 105
The relative importance of the different predictor groups 108
C. Cross-validation studies 111
2. Prediction of performance and adjustment in eight tasks 115
A. Prediction of quantitative performances in eight tasks 115
Measurements of ability as predictors 115
Observational data as predictors 123
Combinations of measurements of ability and observational data as predictors 128
B. Prediction of composite criterion measures 132
Ability measurements as predictors 132
Observational data as predictors 138
Combinations of measurements of ability and observational data as predictors 143
C. Cross-validation studies 146
3. The predictor structure of the ability variables 150
A. The results of factor analyses 150
B. The predictive power of different ability groups 158

Chapter IV. Summarizing Discussion and Conclusions 159
1. Summarizing discussion 159
The starting-points for the investigations 159
Selection and measurements of prediction variables 160
Analysis of the content of the criterion variable 160
Measurements of work performances 161

Observations of working behaviour and composite criterion variables 162
Prediction of prognosis-group and percentage-group criteria 162
Prediction of quantitative performance in eight selected tasks 163
Prediction of composite criterion variables 164
The trustworthiness of the results 166
2. Unanswered questions and needs of further research 168
Identification of valid predictors and analysis of predictor structures 168
The dimensionality of the criterion variable 169
Longitudinal studies and long-range prediction 169
Sex differences 170
Prediction research and the effects of training 174

APPENDICES

Appendix A:1. Correlation matrix for 27 predictor variables. Men ($n$ = 47) 178
Appendix A:2. Correlation matrix for 27 predictor variables. Women ($n$ = 33) 179
Appendix B:1. Sub-scales in the observation of working behaviour 180
Appendix B:2. Sub-scales in a five-degree rating schedule 185
Appendix C. Detailed descriptions of eight jobs 186
1. Pressing eyelets 186
2. Cutting threads in nuts 189
3. Turning holders 193
4. Assembly of brackets 196
5. Assembly of electrical panels 200
6. Making wooden frames 204
7. Embroidering cat motifs 209
8. Packing blocks 213
Appendix D. Control study of the extent of possible transfer effects on the measurement of performance 216
Appendix E. Effects of training over a long period 219

References 225

# Tables


1. Degree of self-support achieved by former pupils of special schools in Sweden 3
2. Comparison between all the employees at 24 out of altogether 29 sheltered workshops in Sweden and the employees of the ALA Foundation, as regards the sex distribution and the distribution between the educable group and the trainable group 27
3. Calculated Reliability coefficients in a number of prediction variables 37
4. Results of ability measurements, expressed as medians, means, standard deviations and ranges 39
5. Mean values and stanine scores in 10 test variables in a normal group of men and in two groups of mentally retarded men at different levels 40

6:1. Number of items, means, standard deviations and estimates of reliability (alphas) in six sub-scales in the observation of working behaviour 49

6:2. Intercorrelations between six sub-scales in the observation of working behaviour 49

7. Number of items, means, standard deviations and estimates of reliability (alphas) in two sub-scales of a five-degree rating schedule 51
8. Correlations between three criterion variables 57
9. The four working groups and their values in the matching variables 63
10. Comparision between the characteristics of the whole experimental group (T) and those of the group which underwent the testing programme (P) 64
11. Average percentages of approved products in eight tasks 64
12. The qualitative development of work performances, expressed as a percentage of the approved products during each of eight consecutive working periods 67
13. The quantitative development of work performances, expressed as the average number of approved products per hour during each of eight consecutive working periods 72
14. The average quantitative performances during each of 4 consecutive days and the corresponding percentage work contribution in relation to the performance expected of the normal worker (according to the calculated MTM value) 73
15. The quantitative measurement of performances, calculated as the average number of approved products per hour during the last 2 working days in eight tasks 76
16. The product-moment coefficients of correlation between quantitative performances during each of the last 2 working days. Total experimental group, $n = 64$ 78
17. Correlation matrix for quantitative measurements of performance in eight tasks and the prognosis-group and percentage-group criteria among men ($n$=40) – below the diagonal–and women ($n$=24)–above the diagonal 79
18. Correlation matrix for quantitative measurements of performance in eight tasks and the prognosis-group and percentage-group criteria. The total experimental group, $n = 64$ 80
19. Factor analysis of quantitative measurements of performance in eight tasks 82
20. Estimates of reliability (alphas) on scale of adjustment to work 86
21. Correlation matrix for two rating criterion variables and eight composite criterion variables among men ($n = 40$), among women ($n = 24$) and in the total group ($n = 64$) 93
22. Factor matrix for 10 criterion variables in the total group ($n = 64$) 95
23. Product-moment correlations between certain composite criterion measures and quantitative performances in selected tasks. Total group. ($n = 64$) 97
24. The correlations of the ability variables with the rating criteria 101
25. Results of stepwise multiple-regression analyses of the prognosis-group criterion 106

26. Results of stepwise multiple-regression analyses of the percentage-group criterion 107
27. Prediction of rating criteria from psychomotor abilities 110
28. Prediction of rating criteria from spatial-inductive and numerical abilities 110
29. Cross-validation coefficients referring to the prognosis-group criterion in the Malmö group 113
30. Results of stepwise multiple-regression analyses of the prognosis-group criterion in two male sub-samples 114
31. Results of stepwise multiple-regression analyses of the percentage-group criterion in two male sub-samples 114
32. Cross-validation coefficients referring to the prognosis-group and percentage-group criteria in the two sub-samples 115
33. The correlations of the ability variables with quantitative performances in eight tasks. Men ($n$ = 40) 116-117
34. The correlations of the ability variables with quantitative performances in eight tasks. Women ($n$ = 24) 118-119
35. Distribution of zero-order correlation coefficients from the point of view of significance 120
36. Prediction of quantitative performances in eight tasks from measurements of ability. Men ($n$ = 40) 122
37. Prediction of quantitative performances in eight tasks from measurements of ability. Women ($n$ = 24) 123
38. The correlations of the observation variables with quantitative performances in eight tasks. Men ($n$ = 40) 124
39. The correlations of the observation variables with quantitative performances in eight tasks. Women ($n$ = 24) 125
40. Prediction of quantitative performances in eight tasks from observations of behaviour. Men ($n$ = 40) 126
41. Prediction of quantitative performances in eight tasks from observations of behaviour. Women ($n$ = 24) 127
42. Prediction of quantitative performances in eight tasks from a combination of measurements of ability and observations of behaviour. Men ($n$ = 40) 129
43. Prediction of quantitative performances in eight tasks from a combination of measurements of ability and observations of behaviour. Women ($n$ = 24) 130
44. The correlations of the ability variables with eight composite criterion measures. Men ($n$ = 40) 133
45. The correlations of the ability variables with eight composite criterion measures. Women ($n$ = 24) 134
46. Prediction of eight composite criterion measures from measurements of ability. Men ($n$ = 40) 136
47. Prediction of eight composite criterion measures from measurements of ability. Women ($n$ = 24) 137
48. The correlations of the observation variables with eight composite criterion measures. Men ($n$ = 40) 138
49. The correlations of the observation variables with eight composite criterion measures. Women ($n$ = 24) 139
50. Prediction of eight composite criterion measures from observations of behaviour. Men ($n$ = 40) 141
51. Prediction of eight composite criterion measures from observations of behaviour. Women ($n$ = 24) 142
52. Prediction of eight composite criterion measures from combinations of measurements of ability and observations of behaviour. Men ($n$ = 40) 144

53. Prediction of eight composite criterion measures from combinations of measurements of ability and observations of behaviour. Women ($n$ = 24) 145
54. Prediction of composite measures of adjustment to work from combinations of measurements of ability and observations of behaviour in men and women 148
55. Prediction of three criteria from combinations of measurements of ability and observations of behaviour in two sub-samples 149
56. Cross-validation coefficients referring to three selected criteria in two sub-samples 150
57. Factor matrix for 21 predictor variables among men ($n$ = 87) 151
58. Factor matrix for 21 predictor variables among women ($n$ = 55) 152
59:1. Ability structure among men ($n$ = 87) 156
59:2. The number of times which ability variables belonging to a certain factor entered into regression equations obtained in altogether 22 stepwise multiple regression analyses of the three different criterion-groups with statistically significant regression coefficients 156
60:1. Ability structure among women ($n$ = 55) 157
60:2. The number of times which ability variables belonging to a certain factor entered into regression equations obtained in altogether 18 stepwise multiple regression analyses of the three different criterion-groups with statistically significant regression coefficients 157
C 1. Number of eyelets pressed per hour 189
C 2. Number of nuts threaded per hour 192
C 3. Number of holders turned per hour 196
C 4. Number of brackets assembled per hour 199
C 5. Number of electrical panels assembled per hour 203
C 6. Number of wooden frames produced per hour 208
C 7. Number of approved cat motifs sewn per hour 212
C 8. Number of cartons packed per hour 215
D 1. Groups in the control study and their values in the matching variables 217
D 2. Measurements of quantitative performances in two groups of experimental subjects for control of possible transfer effects and their magnitude 217
E 1. Number of sponges packed per hour 221
E 2. Percentage work contributions in different tasks 222
E 3. Average quantitative performances in assembly of electrical panels during four weeks in men and women 223

# Figures


| | | |
|---|---|---|
| 1:1. | Age distribution in men belonging to the trainable group | 28 |
| 1:2. | Age distribution in men in the total group | 28 |
| 1:3. | Age distribution in men belonging to the educable group | 28 |
| 2:1. | Age distribution in women belonging to the trainable group | 28 |
| 2:2. | Age distribution in women in the total group | 28 |
| 2:3. | Age distribution in women belonging to the educable group | 28 |
| 3:1. | IQ distribution in men belonging to the trainable group | 29 |
| 3:2. | IQ distribution in men in the total group | 29 |
| 3:3. | IQ distribution in men belonging to the educable group | 29 |
| 4:1. | IQ distribution in women belonging to the trainable group | 29 |
| 4:2. | IQ distribution in women in the total group | 29 |
| 4:3. | IQ distribution in women belonging to the educable group | 29 |
| 5:1. | Intercorrelations in men and women among all other prediction variables and reaction time (•), Tapping (○) and Vocabulary (+) | 43 |
| 5:2. | Intercorrelations in men and women among all other prediction variables and O'Connor test (•), the Leiter Scale (○) and Mark-making (+) | 43 |
| 5:3. | Intercorrelations in men and women among all other prediction variables and Addition, W (•), MTTA (○) and Surface development (+) | 44 |
| 5:4. | Intercorrelations in men and women among all other perediction variables and Block design (•), Threading (○), and Plates (+) | 44 |
| 6. | Distributions of the prognosis-group criterion among men and women | 54 |
| 7. | Distributions of the percentage-group criterion among men and women | 55 |
| 8. | Distributions of the wages-group criterion among men and women | 56 |
| 9. | Sketch of the arrangement for studying transfer effects | 61 |
| 10. | Distributions of the average percentage of approved products in eight tasks among men and women | 69 |
| 11. | Correlation diagram, showing the correlation between quantitative performances (number of approved products) and qualitative performances (percentage of approved products) | 70 |
| 12. | Average quantitative development of performances during 4 days in eight tasks. The performances were evaluated as percentage work contributions in relation to the performances expected of normal workers (according to the calculated MTM values) | 74 |
| 13. | Distributions of average quantitative performances (average of z-values in 8 tasks) in mentally retarded men and women | 85 |
| 14. | Distributions of average ratings of adjustment to work, I (using the check-list form) among men and women | 86 |
| 15. | Distributions of average ratings of adjustment to work, II (using the five-degree rating schedule) among men and women | 87 |
| 16. | Distributions of total stoppage time in minutes among men and women | 89 |
| 17. | Distributions of average ratings of work proficiency, I (using the check-list form) | 89 |
| 18. | Distributions of average ratings of work proficiency, II (using the five-degree rating schedule) | 90 |
| 19. | Distributions of average ratings of quantitative performances in eight tasks | 90 |
| 20. | Correlations of the ability variables with the prognosis-group and percentage-group criteria respectively among men (○) and among women (v) | 102 |

21. Correlations of the ability variables with the prognosis-group criterion (○) and the percentage-group criterion (v) among men and among women 103
22. Proportions of explained criterion variance in the prediction of rating criteria from 3 different combinations of predictors 111
23. Distributions of the prognosis-group criterion among men and women in the cross-validation group 112
24. Proportions of explained criterion variance in the prediction of quantitative performances in eight tasks from different combinations of predictors. Men. ($n$ = 40) 131
25. Proportions of explained criterion variance in the prediction of quantitative performances in eight tasks from different combinations of predictors. Women ($n$ = 24) 131
26. Proportions of explained criterion variance in the prediction of eight composite criterion measures from different combinations of predictors. Men. ($n$ = 40) 146
27. Proportions of explained criterion variance in the prediction of eight composite criterion measures from different combinations of predictors. Women ($n$ = 24) 147
28. Distributions of composite measures of adjustment to work (average of the z-values of three sub-scales) among mentally retarded men and women 148

C 1. Development of performance in the pressing of eyelets 187
C 2. Photo of workplace for pressing eyelets 188
C 3. Sketch plan of workplace for pressing eyelets 188
C 4. Development of performance in threading nuts 190
C 5. Photo of workplace for threading nuts 191
C 6. Sketch plan of workplace for threading nuts 191
C 7. Development of performance in turning holders 194
C 8. Photo of workplace for turning holders 195
C 9. Sketch plan of workplace for turning holders 195
C 10. Development of performance in assembling brackets 197
C 11. Photo of workplace for assembling brackets 198
C 12. Sketch plan of workplace for assembling brackets 198
C 13. Development of performance in assembling electrical panels 201
C 14. Photo of workplace for assembling electrical panels 202
C 15. Sketch plan of workplace for assembling electrical panels 202
C 16. Development of performance in making wooden frames 205
C 17. Photo of workplace for sawing wooden frames 206
C 18. Sketch plan of workplace for sawing wooden frames 206
C 19. Photo of workplace for bench work on wooden frames 207
C 20. Sketch plan of workplace for bench work on wooden frames 207
C 21. Development of performance in embroidering cat motifs 210
C 22. Photo of workplace for embroidering cat motifs 211
C 23. Sketch plan of workplace for embroidering cat motifs 211
C 24. Development of performance in packing blocks 214
C 25. Photo of workplace for packing blocks 214
C 26. Sketch plan of workplace for packing blocks 215

E 1. Sketch plan of workplace in the sponge investigation. Distances of motion given in centimetres 220
E 2. Development of performance among men (upper line) and women (lower line) in assembly of electrical panels during altogether four working weeks 223

# Chapter I

# The Background

It was only after centuries of civilization that it was realized in various parts of the world that society had a responsibility to take care of the mentally retarded. To begin with, it was only children with slight degrees of retardation who were given such care, by being looked after in children's homes, schools and other training centres. Later, this care was extended to include "uneducable" children and mentally retarded adults. Nordström's well-documented historical account (1968, pp. 219–310) makes it quite clear that these events took place in this particular order in Sweden. It is only in modern times that we have begun to realize that the attainment of adulthood as a matter of fact marks an intensification of the problems of adjustment for the mentally retarded and that at this particular time they have an especially great need of society's support. It will probably be a long time before people are more generally prepared to draw the consequences of this realization by making sufficiently comprehensive, varied and individually oriented efforts at practical habilitation, but the beginnings of this can be discerned in different parts of the world. (On the current situation in Sweden, see Grunewald 1969.)

## 1. Society's Care for the Mentally Retarded

When society is to take care of mentally retarded adults, questions concerning vocational guidance, vocational training and adjustment to work will naturally be in the focus of attention. The most important aim of any action taken must be to develop, on the basis of individual aptitudes, the capacity for independent living in as many of them as possible. The mentally retarded person who can earn his livelihood gains respect in his own eyes and in the eyes of others and also relieves society of considerable expense. This double and incontrovertible observation has also long been used to justify the argument that society ought to make resources available for the judicious habilitation of the mentally retarded. For example, C.U. Sondén, the Swedish psychiatrist, used a similar formulation in the pamphlet entitled "Om idioters uppfostran och undervisning" (The Upbringing and Instruction of Idiots), which he published as early as 1857 (cf. Nordström 1968, p. 229). A good century later, Wolfensberger gave expression

to the same basic view in a more differentiated way when he formulated his "vocational creed" (Wolfensberger 1967, pp. 232–3).

Long before the official social-welfare organizations were prepared to take on responsibility for facilitating the adjustment of the mentally retarded to work and social life, many private persons made important efforts in creating a favourable climate of public opinion and in taking initiatives. In modern times, the parents' organizations have often been able to play an important part in this respect. In Sweden, one example of this is the ALA Foundation, which was founded in the mid 1960s by the Swedish Association for Mentally Retarded Children, in collaboration with the Swedish Red Cross and the Swedish Boy Scouts Association.

The name ALA was formed from the initial letters in the Swedish phrase *anpassning till liv och arbete* (adjustment to life and work), but it simultaneously forms the Latin word for "wing" and in this way symbolizes in a double sense the Foundation's endeavour to make mentally retarded young people capable of independent living by giving them all-round training. More concretely, the task was stated to be "to establish in Uppsala a sheltered workshop with an associated home for boarders, to organize adult education, to arrange leisure programmes, to give staff training, to disseminate information and to carry on scientific research with reference to these activities".

The programme formulated in this way was carried out in a few years of intensive work, during which valuable experience was gained. Here I shall only mention the Foundation's very conscious endeavours to supplement the practical work of habilitation with interdisciplinary research. For this purpose, it engaged a comprehensive group of research advisers, including qualified representatives of several disciplines in medicine and behavioural science, and also engaged on a full-time basis a medical research worker, a psychological research worker and a welfare officer. Through the discussions and consultations held in this group, there eventually developed an initial research programme, in which the aim was stated to be to study the habilitation of mentally retarded adults in different subsidiary investigations and to analyse the essential conditions for it. In an application for grants, three sets of conditions were distinguished, as follows:

(1) The inherited and acquired potential in mentally retarded persons for adjustment to life and work.

(2) The methodological design of different measures of habilitation, such as work training, home training, social-function exercises and various forms of supporting and stimulating therapy.

(3) Such important environmental factors as the attitudes of neighbours, friends and relations and the current structure of the labour market.

The studies presented in the following pages form a part of this research programme and are chiefly connected, as may easily be ascertained, with the

first-mentioned set of conditions for habilitation, namely, the potential in mentally retarded persons for adjustment to life and work.

## 2. The Potential for Adjustment to Work and the Need of Individual Guidance

One question of fundamental importance in this connection is, of course, what chances the mentally retarded person has of fulfilling in an acceptable manner the functions of an employee in the community. For decades, researchers in different countries have made repeated follow-up investigations—using gradually improved methods—for the purpose of finding an answer to this question. Thus, in Sweden, similar investigations were carried out on the initiative of the National Board of Education in two different years, viz. 1949 and 1962 (Jungner 1955, Wessman and Gustin 1964). As regards the degree of self-support achieved, Wessman and Gustin summarized the results as shown in Table 1.

*Table 1.* Degree of self-support achieved by former pupils of special schools in Sweden.

| Degree of self-support | Men | | Women | |
|---|---|---|---|---|
| | 1949 | 1962 | 1949 | 1962 |
| Completely self-supporting | 38.7 | 42.7 | 26.7 | 21.8 |
| Partly self-supporting | 31.1 | 20.9 | 32.0 | 24.0 |
| Not self-supporting | 30.2 | 34.4 | 41.3 | 47.5 |
| No information | – | 2.0 | – | 6.7 |

The first group studied consisted of all the pupils who completed their educations in the special school during the period 1936—45. The second investigation included all the pupils who were born in the period 1935—39, who had been admitted to a special school and who had completed their educations there. Thus, in both cases, it was a question of population studies with a response frequency of 100 % or nearly 100 %, a circumstance which makes the figures particularly interesting. Table 1 shows that about two-thirds of the men and about half of the women were found to be wholly or partly self-supporting.

However, most of the follow-up investigations of this kind have been carried out in the English-speaking countries and there are several comprehensive annotated surveys of them (Goldstein 1964, Tizard 1965, Kokaska 1968, Cobb 1969). Goldstein shows how the positive results which Fernald achieved in such investigations as early as 1919 helped to clear away the basic attitude of

pessimism that had previously been predominant. Fernald and his successors found—with comparatively large variations in the figures reported—that most mentally retarded persons possessed the ability to achieve and retain an acceptable level of adjustment to work and to social life, though the great majority of them had to be content with unskilled work.

Of particularly great interest in the long series of follow-up investigations are the studies in which the authors have observed the behavioural development of mentally retarded people for long periods of time and made comparisons with control groups of people of normal intelligence. Cobb (1969) reports and analyses thoroughly two such investigations, one made by R. J. Kennedy in Connecticut in the period 1944–60 and the other by Baller, Charles and Miller in Nebraska in the period 1935–64. Both the authors themselves and Cobb consider that these studies confirm not only that most mildly retarded adults have the ability to achieve a satisfactory adjustment to social life but also that, as the years go by, they show an increasing tendency in this respect. An increasing number of them, in fact, relinquish their status as mentally retarded persons and "disappear into the general population".

In several accounts of such follow-up investigations, the authors are anxious to emphasize how complicated the problem actually is. Whether or not the mentally retarded person succeeds in achieving an acceptable level of adjustment to work depends not only on his own potential but to a very large extent also on such external factors as the prevailing situation on the labour market and the attitudes of the management and the other employees. Comparisons between follow-up investigations from different periods confirm, for example, that the percentage of mentally retarded persons who could be described as being in "gainful employment" was considerably lower during the period of economic depression at the beginning of the 1930s. The decrease demonstrated during this period was also found to be much more obvious in this group than in a simultaneously studied control group of persons of normal intelligence (Goldstein, pp. 239 ff.).

Other factors which are also of essential importance for the mentally retarded person's adjustment to work are changes in the structure of commercial and industrial life and changes in production technique, particularly those due to increased mechanization and automation. With reference to the American labour-market statistics, Goldstein shows that "the number of jobs usually identified with the retarded is on the decrease and that competition for these jobs will become more keen" (p. 251). Obviously, such developments increase the difficulties for the mentally retarded and give Goldstein good reason to stress the need to give them individual guidance in connection with a purposeful education and training.

However, the question of the mentally retarded person's prospects of adjusting himself to work and social life is complicated also by the fact that it actually

refers to a highly heterogeneous group. Goldstein's account of the follow-up investigations is concerned throughout with the "higher-grade mentally retarded", whether they were later institutionalized or not. Thus, in reality the individuals studied formed a positive selection from among the mentally retarded. Kokaska may be said to have gone a step further in the same direction when he restricted himself to former pupils in special classes with a variation of IQ between 50 and 85. Together with the more favourable conditions on the labour market, this circumstance would seem to explain why Kokaska finds reasons for coming to more positive conclusions than Goldstein, as regards the working capacity of the mentally retarded. Whereas Goldstein observed, by way of summary, that "the retarded cluster in the unskilled category", Kokaska found that several follow-up investigations "indicate that the subjects were able, over several years, to progress from unskilled to semi-skilled and from semi-skilled to skilled areas of work".

However, it must not be forgotten that today society's care also embraces groups of more seriously retarded persons who were often previously described as uneducable. Their capacity for adjustment to work has hitherto received far less attention. However, some investigations have been carried out in recent years in order to shed light on this question also, as is clear from the accounts given by Tizard (1965), Clarke and Clarke (1965) and Guskin and Spicker (1968). Tizard refers chiefly to an English report from Birmingham concerning a good 1000 individuals who as children had been excluded from school as uneducable, and to two American investigations carried out by Saenger, comprising a representative sample of 520 "adults with a moderate level of retardation" who as children had been pupils in special classes in New York City. In both these cases, it was found that the average ability to adjust was greater than expected and that a not inconsiderable number could support themselves. At the time of Saenger's investigation, 20 % of the individuals whom he studied were institutionalized. Most of the rest were well accepted in their homes and in their immediate surroundings. Certain deficiencies in social adjustment were noted. For example, they had few friends and were often lonely and unemployed. However, no less than 27 % of them were working either full-time or part-time and Tizard considers that he is justified in interpreting the results as follows: "These studies suggest that, while we still have no clear conception of the potentialities of adult imbeciles, at least 10 per cent and probably more may be capable of earning a living in the community. Others can work in sheltered emptoyment, given suitable conditions" (p. 488).

Cobb (pp. 51 ff.) reports Saenger's investigations in detail and emphasizes especially that the surprisingly good results were achieved without the community supplying any other supporting services than institutional care. With reference to observations reported later from the United States and England, he considers

that we may expect "much higher levels and frequencies of adaptation by the adult moderately retarded through intensive social-educational and occupational services".

We meet with a similar view of the developability and capacity for social adjustment of these groups in Stevens (1964) in connection with his presentation of and commentary on the classification of the mentally retarded in four different degrees of retardation to which the American Association on Mental Deficiency has given official sanction. My own experience of habilitation work for 5 years at the ALA Foundation points in the same direction. If also a comparatively small minority of moderately retarded adults may be expected to become self-supporting and socially independent, all those who belong to this group may profit from systematic training specially arranged for them and thereby achieve a higher level of adjustment in several respects. Under these circumstances, it is evident that the individually designed guidance which was found necessary for mildly retarded adults will be at least equally necessary for this less well-endowed group.

## 3. Methodological Problems in Previous Prediction Studies

As the first stage in the necessary individual guidance of mentally retarded persons, one has to analyse the individual's qualifications, as regards training, placement in work and choice of occupation. In this connection, the chief problem is not to observe and measure a number of different individual characteristics but just to gain an insight into their prognostic signification. As Windle (1962) and, long before him, Goddard (1909) pointed out, it is extremely important from several points of view to be able to carry out this task in a reliable fashion. Sometimes the situation is such that one has to select a limited number of individuals from a whole population for a certain treatment or a certain training, and in that case it is important to choose those who are most receptive to the treatment or who can profit most from the training. Perhaps even more important is the view—expounded by several researchers—that knowledge of the prognostic value of different variables may guide the design of basic training and educational programmes. It is of essential importance that endeavours should be made in this connection, above all, to stimulate, develop and supplement the variables which show the closest correlation with the adaptability and efficiency demonstrated later. As Cobb (1969) convincingly shows, the need of prognostic information increases at the same rate as one abandons the typical selection thinking and tries to give individually oriented guidance, "which aims not so

much at the highest ratio of successes to failures as at the maximum habilitation of each individual who needs assistance" (p. 5).

It is insights of this kind that have for decades impelled administrators and researchers to collect data and to try to establish the prognostic values of different variables. Several summaries and evaluations of such investigations have been published (Tizard and O'Connor 1950*a* and *b*, Shafter 1957, Windle 1962, Patterson 1964, Wolfensberger 1967, Cobb 1966, 1969). In this connection, methodological problems have been extremely topical.

Windle's survey is comprehensive but suffers from the limitation that it only describes investigations on mentally retarded persons who are institutionalized. From that point of view, it is best supplemented by Wolfensberger, who in his turn pays most attention to similar studies of mentally retarded people in communities. Both these authors are agreed that the great majority of the prognostic investigations that they have examined show palpable deficiencies from the methodological point of view and that on this account they provide only very limited support for reliable conclusions. Windle points out in particular that "prediction in mental subnormality is a much less mature field of investigation than prediction in parole from penal institutions or in the outcome of mental illness" (p. 133).

The deficiencies to which Windle and Wolfensberger call attention refer to all stages of the methodological procedure applicable in connection with prognostic investigations. Thus, as regards the selection of experimental subjects, Windle points out that "availability" has often been "the prime determiner of the population studied" (p. 27). This has meant, *inter alia,* that the requirements for fairly homogeneous and sufficiently large groups have seldom been met. In addition, "placement biases" have often been allowed to determine the compositions of the groups studied (p. 28). Prognostic characteristics or predictor variables have often been chosen without sufficient knowledge and the choice has seldom been preceded by the formation of any hypothesis. As a rule, the criteria applied have been vague and consequently unreliable. In most of the investigations, the authors have used either an administrative classification, such as "Discharged" and "Not discharged", or clinical ratings of the degree of social adjustment achieved. In both cases it may be assumed that the practice and the ground of assessment varied greatly from one institution to another. In this connection, it is important to note that the authors often allowed the follow-up period to vary within comparatively wide limits in the groups studied. Finally, Windle and Wolfensberger point out that the authors of the works reviewed have not paid sufficient attention to the problems of significance and that they have practically always exempted themselves from the requirement to make cross-validation analyses, which is an essential requirement from the methodological point of view.

Thus, though the numerous prognostic investigations which Windle, Wolfensberger and others have critically examined have not to any marked extent resulted in reliable and practically applicable knowledge, they have the merit that they can be made the starting-points for further research with improved methods. Windle emphasizes this and finds it particularly important that authors should in future concentrate on analysing the prognostic relevance of well-demarcated and well-defined variables belonging to four different fields, namely, (1) diagnostic classifications, (2) abilities and ability patterns, (3) personality characteristics, and (4) forms of treatment (p. 137).

Wolfensberger also devotes his attention chiefly to the choice of predictor variables, in analysing the possibilities of further research. However, his view is that hitherto authors have concentrated too much on variables related to the individual and have neglected "the variables inherent in the training and placement process or those associated with the placement situation" (p. 242). For the rest, he discusses in comparative detail "work-sample procedure" and considers that it has not yielded results commensurate with the money which has been spent on it but that with a modified design it may be a better prognostic instrument. Like certain other researchers, he also puts forward the view that one should start from the "improvement rate" or an equivalent variable, instead of test scores in the traditional sense. Finally, he emphasizes that "social, background and observational data", as well as the occurrence and degree of various handicaps, should be tested as predictor variables to a greater extent than hitherto.

Patterson (1964) also presents similar critical views regarding "work-sample procedure". Moreover he emphasizes more strongly than the others that one should not neglect personality factors, for example, persistence and attitude to different types of tasks, in the search for essential predictor variables. As, at the same time, ordinary methods of personality analysis, such as Rorschach, TAT, MMPI, etc., are not very suitable for the mentally retarded, one should instead endeavour to arrange situations which allow of as reliable observations as possible to be made of such personality traits. Like several other researchers Patterson emphasizes finally that further efforts should be made "to isolate and measure the basic components of manual performance" (p. 146).

## 4. Modern Lines of Methodological Development

### A. The combination of prediction variables

Most of the prediction studies carried out up to the middle of this century were aimed at analysing, in proper order, the relation between a single predictor variable and a dichotomous criterion. However, Windle described in 1962 some attempts to combine prediction measurements in various ways, in order to

produce a more accurate prediction, and during the last 10 years similar procedures have been applied in an increasing number of studies.

O'Connor and Tizard (1951) present an early example of this multivariate approach. They attempted to predict the degree of employability of 104 mildly retarded boys with a mean age of 20.9 years on the basis of 13 prediction variables. After applying the square-root method of selecting the variable for multiple regression, multiple correlations were obtained with the criterion which amounted to .67. In this study, the satisfactory predictors appeared to be the Rail-walking Test, the Body-sway Test, the Progressive Matrices Test, two combined manual-dexterity tests and a four-point rating of instability. The authors are of the opinion that it should be possible to use the results in practical selection situations "with reasonable hope of success", but at the same time they point out that "no undue emphasis is placed on the significance of correlational findings of this order".

Cobb (1969) presents a detailed and thorough evaluation of a number of prediction studies made during the 1960s. All these studies are also examples of the endeavour to achieve in different ways a combination of a limited number of variables with the maximum predictive power. In the following pages, I shall give accounts of some of these studies.

Taylor (cf. Cobb, pp. 59 ff.), who studied an experimental group of 74 subjects, started from a total of 44 scores, which he classified as tests of intelligence, tests of dexterity, tests of maladaptive personality traits, tests of educational attainment and tests of specialized or mixed functions. He first determined the predictive value of these indices by correlating them with certain criteria variables. With the aid of this information, he then selected four tests as being the most effective predictors, namely, (1) WAIS, Digit Symbol, (2) Bender Gestalt, (3) Purdue Pegboard Assembly, and (4) Wrap-o-Pack (a specially designed experimental task). Multiple correlations were calculated between this combination of tests and three criterion variables and came to .47, .67 and .70 respectively.

Ferguson and Pinkard (cf. Cobb, pp. 73 ff.) in their study developed a set of eight prediction variables, which were summarily described as the Vocational Capacity Scale. This included three rating scales, referring to working habits, physical capacity and general state of health, and five tests, which were intended to test motivational persistence, compliance with verbal directions, social maturity, manual skills and arithmetic. The correlations of these variables with the criterion were calculated in the form of contingency coefficients and chi-square values. With the aid of this information, a weighting procedure was carried out, before the different variable scores were finally added up to give a total score, which proved to have a certain capacity to discriminate between the authors' groups of criteria.

A few years later, the Vocational Capacity Scale was re-validated by Dayan

(cf. Cobb, pp. 78 ff.). However, Dayan used a different method of combining the sub-scales. He converted the individual variable values into standard *z* scores and then added them up. He was able to show clearly significant differences between the mean values of this composite standard score in his three criterion groups, which he designated "Competitively employed", "Sheltered employed" and "Non-employed". However, he was not content with this but then converted the criterion into an 11-point scale of job difficulty or complexity. The calculated correlations between this criterion and the sub-scales in the Vocational Capacity Scale varied between .36 and .71. When, finally, Dayan also calculated the multiple correlation between six of the sub-scales and the criterion, this correlation was found to amount to 0.79.

Schulman (cf. Cobb, pp. 86 ff.) started with about 10 predictor variables. With the guidance of a rotated varimax factor analysis, he considered that he could classify these variables in one of two groups. The "competence group" consisted of "measures of intelligence, production, manual dexterity, ratings of general work behavior and, to a lesser extent, social maturity, which seemed to covary with each other positively and systematically". The "self-concept" group could not be identified equally consistently in the three analyses which were carried out. However, in this group he classified Projected Drawings, the Guthrie Self-attitude Scale, certain interview ratings and the stated levels of aspiration in connection with the performance of certain tasks. The criterion was designated "Employability" and consisted of a combination of a production score and diagnostic work ratings. The correlation between this and the predictor variables was established partly by a dichotomous analysis of variance between high and low employability groups and partly by calculating product-moment correlations and making a stepwise-deletion, multiple-regression analysis. In both cases, the results seemed to confirm that there were positive correlations between employability and the competency cluster, while the correlation between the criterion and the self-concept variables was non-existent or, in one or two cases, perhaps negative. The Purdue Pegboard emerged as by far the best prediction variable, in that its correlation with the criterion came to .70. The stepwise, multiple-regression analysis showed that, by adding a further six of the prediction variables, it was possible to obtain a multiple correlation of .76.

The examples reported above of the endeavour to combine different prediction variables will undoubtedly stimulate researchers to make further efforts in the same direction. Though no definitive results can be said to be available until replications and real cross-validations have been carried out on a larger scale, it seems to be evident that the accuracy of the prediction can be increased by allowing several independent variables to contribute to it, either by adding together scores on sub-scales or by different forms of multiple-regression analysis.

However, the extent to which a battery of predictive tests serves its intended

purpose is determined chiefly by the nature of the sub-variables included. If we consider the above-mentioned examples of studies from that point of view, we find that various psychomotor skills consistently display the highest correlations with the criteria. However, tests referrring to generally cognitive functions and ratings of social maturity and behaviour in work situations also contribute to the prediction. Whether personality factors of other kinds can serve as predictors is much more obscure.

**B. The analysis of the problems of criteria**

In parallel with the endeavour to combine different prediction variables, in order thereby to achieve a predictor structure which may explain the greatest possible variance in a criterion, one finds in certain studies from later times an equally conscious endeavour to similarly analyse and elucidate the structure of the criterion variable.

Criteria in the form of administrative classifications such as "Discharged" and "Not discharged", which have been used in a large number of investigations, still have, of course, a definite interest from special points of view. They are applicable if one wishes to study institutional processes. for example, the effects of different forms of treatment in an institution (Windle, p. 29). But the compositions of the clienteles differ from one institution to another, which makes comparisons between different studies very difficult and limits the possibilities of reaching any general conclusions. In addition, there is the important fact that the applied discharge strategies, which presumably vary from institution to institution, "result in a distortion of the evidence regarding the determinants of successful community adjustment" (Cobb, p. 56). This unavoidable complication has been given detailed attention in modern Swedish studies in this field (Kebbon 1969. Holmlund *et al.* 1971). Finally, it should be remembered that, on account of their nature, these criteria are only applicable in studies of mentally retarded persons who are institutionalized. This is a serious limitation at a time when increasingly deliberate endeavours are being made to normalize the environment in which mentally retarded children and young people grow up and not to place them in institutions except in cases of necessity.

Thus, if we reject, on the whole, administrative classifications of the kind mentioned above as suitable criteria, it remains to try and define the meaning of successful social adjustment and unsuccessful social adjustment and to find methods of rating or, best of all, measuring this variable. This is also exactly the procedure that has been applied, in a more or less sophisticated fashion, in many studies.

It is possible to mention several examples of authors who in this connection have simply confronted judges with the task of characterizing in two or more steps the social adjustment achieved as in a general sense good or bad. It is evi-

dent that this is an unsatisfactory method, chiefly because social adjustment is an extraordinarily complex variable. One judge may keep in view an aspect of adjustment which another may entirely disregard.

One may get a concrete idea of how complex a variable social adjustment in fact is by recalling how the authors of the above-mentioned follow-up studies set about assessing, in a comprehensive way, the degree of social adjustment achieved by the mentally retarded. A good example is the Kennedy study, in which the author at the follow-up tried to determine the status of the subject population in the following respects: marital adjustment, offspring (the reproduction rate, IQ, school performance), economic status, work performance, dependency and relief, anti-social behaviour, social participation and leisure-time activities (Cobb, pp.28ff.).

In the more analytically oriented prediction studies, it is only in recent years that authors have made more serious attempts to do justice to the complex or heterogeneous nature of this criterion. I shall consider two examples of this in greater detail.

In Stephens's study entitled *The Success of Young Adult Male Retardates,* which is one of those evaluated by Cobb, it is obvious that the author "assumes from the outset an extreme equivocation in the concept of successful adjustment and undertakes to bring this ambiguity into some kind of order through systematic analysis" (Cobb, p. 97). His procedure was as follows. With the aid of a panel of experts, he found that no less than 141 variables were needed to cover the whole adjustment process. He considered at the outset that these variables could be classified in three categories of behaviour, namely, "personal success", "socio-civic success" and "vocational success". Measurements of these variables were obtained by designing questionnaires for use in structured interviews with the experimental subjects, their parents or other relations, and employers and supervisors. Certain data were subjected to factor analysis, in which 17 criterion factors were obtained as a result. To these were added 61 non-reducible dichotomous variables.

In order to be able to predict a criterion of such a complex structure, Stephens designed a multidimensional battery of predictors. This included, from the beginning, no less than 78 variables, of which only a few belonged to the generally cognitive and psychomotor sphere, while the bulk of them were personality measures in the form of interview ratings, self-report inventories and also certain projective measures. By factor analysis, these 78 variables were reduced to 21 factors. The correlations between these factors and the criterion variables were studied by multiple-regression analysis. A large number of statistically significant correlations could also be demonstrated, but their interpretation presented considerable difficulties and Cobb observes very rightly that "the effect of the original instrument has been so diluted through the successive steps that no

meaningful statement of its direct contribution to the variance of the criterion factor can be made" (p.109). I also consider that there is reason to emphasize the justice of Cobb's doubt that satisfactory reliability had been achieved in the many ratings, as well as his misgiving that the halo effect had been allowed to play too important a part. I am therefore inclined to regard Stephens's study primarily as an illustrative example of the considerable difficulties which one encounters as regards measuring technique if one sets one's mind on thoroughly analysing the structure of the adjustment process in mentally retarded persons in all its complexity and on predicting it.

A recently published small-scale study by Rosen *et al.* (1970) also testifies to these difficulties. These authors started from the need of uniform criteria and considered that these criteria should cover "as many facets of social and vocational experience as possible". In a study of 65 individuals recently discharged from an institution, they used 22 "criterion measures derived from personal interviews, ratings and on-the-spot evaluations of the subjects' social, vocational and economic adjustment in the community". The correlations obtained between these measures and 29 prediction variables are stated to be "relatively low". During the next two years, the authors managed to carry out two replications (Rosen *et al.* 1972), which, however, yielded a negative result. The latter is, of course, of great interest. The authors are of the opinion that the explanation of the negative result probably lies in the wrong choice of predictor variables. However, the question is whether the reasons must not also be sought in the difficulties of reliably measuring so many and such specific aspects of the criterion.

In several other studies, the authors have—more or less deliberately—refrained from designing the criterion variable so that it covers the process of adjustment as a whole and have instead restricted it to refer to working capacity and adjustment to work. One example of this is the study by O'Connor and Tizard (1951), which I described earlier from other points of view (cf. above p. 9). After these authors had collected data in the prediction variables, they endeavoured to get as many as possible of the 104 experimental subjects employed among normal workers, either on building work or in factories on daily licence. After this employment had gone on for nine months, they had a social worker rate them all in accordance with the following seven-degree scale (p. 208):

| | |
|---|---|
| Conspicuous all-round success on daily licence | 7 |
| Settled down well in first job on daily licence | 6 |
| Satisfactory on daily licence after no more than two re-placements | 5 |
| Settled daily licence after several trials | 4 |
| Failed consistently on daily licence | 3 |

| | |
|---|---|
| Can work in institutional conditions but in the judgement of the Chief Male Nurse and Social Worker is unlikely to be considered for licence | 2 |
| Was unsuccessful in institution workshop | 1 |

It is obvious that this was a question of ratings which could be based on prolonged observation under conditions which were relatively comparable from individual to individual and which had to a considerable extent the character of empirical trials. The reliability of such a criterion may be assumed *a priori* to be satisfactory. The criterion also has the merit of being highly relevant. What one feels is most urgent from the point of view of practical habilitation is really to know whether the mentally retarded who are the objects of this care can cope with a job on the open market or otherwise whether they are amenable to work training in more sheltered circumstances.

In the work of Ferguson and Pinkard and again in that of Dayan (cf. above, p. 9), we meet with a classification which in all essentials tallies with the seven-degree scale used by O'Connor and Tizard. Ferguson and Pinkard carried out their study at the MacDonald Training Center, which at that time admitted mentally retarded persons aged 15—30 and passed them through a programme which comprised four phases, namely, (1) evaluation (six weeks), (2) training (from six months to two years), (3) work in a sheltered workshop (for a varying period), and (4) external placement. However, all the clients could not achieve the ultimate aim of external placement but had to continue to work in a sheltered workshop or were referred to some form of non-economic, day-care type of management. Thus, as the final result of the whole procedure, the clients were referred to one of three terminal categories, designated "competitive employment", "sheltered employment" and "work activity" (cf. Cobb, p. 73). Dayan proceeded in a similar fashion but carried out a further differentiation by starting from the degrees of difficulty of the various tasks. This made his criterion variable an 11-degree rating scale. This also may be said to be quite empirically based and is at the same time suitable from the point of view of practical habilitation.

Thus, though the criterion variables used by O'Connor and Tizard, Ferguson and Pinkard, and Dayan have obvious merits, they are still characterized to a certain extent by the defects which are always inherent in subjective evaluation. It is therefore not surprising that some researchers have tried different approaches and that they have also tried to measure the work outputs of mentally retarded persons, in order to arrive at wholly objective criteria by this procedure. Examples of this approach will be found in the works of Schulman (cf. above, p. 10) and Elkin (1967, 1968).

Schulman had his experimental subjects work for three hours a day for five days a week on several different tasks. The production was measured day by day

and the average output during the week was calculated as a percentage of the values expected from normal adult workers. In addition, "the appropriateness of the worker's behavior on the job, other than production" was observed and rated. This was done with reference to an 11-point diagnostic scale, comprising 25 items. The production score and the work ratings were both converted into standard scores with a mean value of 50 and a standard deviation of 10 and were then added together to form a criterion of employability. I have already described above how Schulman found correlations between this criterion and some of the predictor variables.

The interesting things in Schulman's study from the criterion point of view are thus, on the one hand, that he attempts to measure the work outputs of his experimental subjects and, on the other, that he combines the measurement of performance with a rating of the appropriateness of the behaviour at work. It is also interesting to note that he tries to validate the criterion used by making it serve as predictor variable in a follow-up comparison of the status five years after the beginning of the study. In order to determine the degree of successful social adjustment—which in this case was the criterion variable—Schulman started from such characteristics as "presently holding a stable job", "history of employment instability", etc. He was able to demonstrate in a two-by-two contingency table a significant correlation between the variables, but the comparison referred to only 23 individuals and one can only concur with Cobb's refusal (op.cit., p. 96) to accept this as a satisfactory validation.

Elkin (1967) stated that one of the aims of his studies was "objectifying criteria for measuring success on the job". He designed four comparatively simple experimental tasks, which were designated "color sorting", "form sorting", "hand assembly" and "tool assembly". After instruction and preparatory training "up to a predetermined criterion", the performances of the experimental subjects at each task were measured for 10 minutes. The measurements obtained in this way were regarded as work-sample tests and were included in the predictor variables in the study. Then came the criterion determinations. The experimental subjects had to work on each of the experimental tasks for 45 minutes on six consecutive days. The measure of performance was calculated as the number of units correctly completed during the last five days. By this procedure, Elkin considered that he could obtain "an objective and continuous measure of performance on each of the jobs".

In the first of the investigations which Elkin reports, 58 "trainable residents" took part (30 males and 28 females). Their ages varied between 15 and 52 years and their IQs between 20 and 55. Altogether 18 prediction variables were used, consisting of the above-mentioned work-sample tests, together with intellectual, psychomotor and biographical measures. Elkin found significant and high correlations between the four different criterion measures and evidently considered

that he could therefore put them together to form "total work scores". He also found significant correlations between the great majority of prediction variables and these total work scores. Only one vocabulary test and the three biographical variables differed from this pattern. As, in addition, Elkin was able to record marked intercorrelations between the successful predictors, he was of the opinion that the results "supported the hypothesis that a general ability factor encompassing intellectual and psychomotor skills is related to job success" (pp. 579–580).

In a subsequent study, Elkin applied similar experimental conditions but this time had "educable residents" as experimental subjects (29 males and 30 females). Their IQs varied between 50 and 75. In this case, the criterion was determined in a somewhat more summary fashion. Only one of the experimental tasks was used, namely, tool assembly, and the effective working time was restricted to $4\frac{1}{2}$ hours on one and the same working day. The number of predictor variables was increased slightly, but there was no change as regarded their general composition. On the whole, the results of the earlier study with trainable residents were confirmed, though the correlations between the predictor variable and the criterion were somewhat lower on this occasion.

Of special interest from the criterion point of view was the fact that Elkin endeavoured to establish the validity of the experimental task. In connection with the experiment just described, the work which 15 of the experimental subjects performed in the institution's sewing-room was rated. When afterwards this rating was correlated with the performance on the experimental task, the correlation proved to be of a significant order ($r$=.60, $p$ <.02) and Elkin considered, not without reason, that this result "suggested a broader application of research findings from such settings than might be expected".

However, Elkin was less successful when in a third study he wished to predict the suitability of 18 mentally retarded women for non-institutional domestic jobs on the basis of nine prediction variables, which included intellectual, psychomotor and work-sample measures. Only the O'Connor Finger-dexterity Test showed a significant correlation with the rating criterion. Elkin considered that this negative result may have been determined by the discharge strategy used in the actual institution, as well as by deficiencies in the scales used in rating the performance on the domestic jobs.

In the light of the attempts briefly described in the foregoing pages to master the criterion problems in prediction studies of mentally retarded persons, the views and conclusions on this question may now be summarized as follows. One is justified in starting from the situation of the institutionalized mentally retarded person and in applying criteria in the form of administrative classifications such as "Discharged" and "Not discharged" only if one chiefly wishes to shed light on institutional conditions and processes. In other cases, one is reduced to deter-

mining what is meant by successful and unsuccessful social adjustment and to trying to develop methods of rating or measuring this variable. The difficulties which one encounters in this connection are due, *inter alia,* to the very complex structure of the variable. The attempts made to analyse it by factor analysis show that one has to allow for a large number of sub-variables and that several of them cannot yet be measured with satisfactory reliability. In the present situation, a more practicable way would therefore seem to be to restrict the criterion variable to refer to work performance and adjustment to work. From the point of view of practical habilitation, this means that highly relevant factors are taken into account and furthermore the situation becomes more manageable from the point of view of measuring technique. Work performances and adjustment to work can be directly observed under controlled conditions and measured both quantitatively and qualitatively. The studies with this orientation described above have also yielded promising results.

**C. Prediction studies among the moderately retarded**

Elkin's study is an example of the fact that in recent years researchers have begun to investigate the possibilities of predicting work performance and adjustment to work also among the moderately retarded. This endeavour has been a natural consequence of the greater understanding of the developability and the surprisingly good social adaptability of this group (cf. above, pp. 5–6). Elkin's group of "trainable residents" had IQs which varied between 20 and 55 and thus included also some severely retarded persons. We have already seen that the correlations in this group between predictors and criteria proved to be in fact somewhat higher than in the group of "educable residents" and that the possibilities of prediction in this group must therefore be described as good. The results of other investigations carried out by Wagner and Hawver (1965) and McCarron and Dial (1972) point in the same direction. The former investigation was concerned with 27 mentally retarded subjects at a sheltered workshop who had IQs of less than 50. Seven different tests in the sphere of psychomotor function were used as prediction variables, together with IQs on the Stanford-Binet scale. The criterion was an overall ranking of workshop success by the chief instructor. It turned out that the calculated ranking correlations between the predictors and this criterion variable were significant in all cases and that several of them were remarkably high.

McCarron and Dial achieved to a large extent parallel results. Their experimental group consisted of 21 mentally retarded workers with a mean IQ of 49. In addition to psychomotor proficiency tests, the prediction variables also included "a quantified neurological assessment of fine and gross motor coordination" and estimates of verbal intelligence. The criteria used were, on the one hand, the San Francisco Vocational Competence Scale and, on the other, a

productivity score in a sample work task. In this investigation also, all the predictors were found to be significantly correlated to the criteria and the multiple-regression analyses carried out resulted in multiple correlation coefficients which came to .83 and .84, respectively.

Thus, it appears to be well attested that, with the aid chiefly of psychomotor test variables, there are comparatively good prospects of predicting the work performances and the adjustment to work of moderately retarded persons and, to a smaller extent, also of severely retarded persons. In view of this, one is inclined to agree with the statement of Wagner and Hawver (p. 690) that "the use of psychological tests with mental defectives in work settings warrants renewed interest".

**D. Cross-validations and replications**

In one prediction study after another, the authors have emphasized the need of cross-validations and replications. They have often done so with reference to the fact that the study presented has, for various reasons, comprised a comparatively small sample. They have also pointed out that the absence of such control studies is a particularly serious deficiency in multiple correlation studies, "since even more shrinkage is expected with optimal combination of variables than is the case with single variables" (Windle, p. 131).

In spite of the fact that research workers have clearly realized the need of control studies, I have not encountered more than a single real cross-validation, namely, that reported by O'Connor and Tizard (1956) nearly 20 years ago. The result of this was almost depressing, as the authors were compelled to report that the multiple correlation with the criterion declined from about .60 or .50 to .25 when they applied the regression weights obtained in the original sample to new data. Both Windle (p. 129) and Cobb (p. 11) have pointed out, however, that the authors did not use identical criterion variables in the two investigations and that this circumstance may have contributed to the dissimilarity in the findings.

Not many replications, in the sense of studies repeated under the same conditions, have been carried out either, but from this point of view the situation may be said to be somewhat more favourable. I mentioned above that Rosen *et al.* (1972) reported two cases with negative results. However, this is of limited significance, as it was a question of an investigation in which the authors attempted to measure several aspects of a complex criterion variable by comparatively untested and probably not very reliable rating methods. No one has yet succeeded in showing that we can make predictions of that kind with satisfactory accuracy. As I have already pointed out in connection with the discussion of criteria, the situation is more favourable when we restrict the criterion variable to refer to work performances and adjustment to work. Dayan's studies (the Pinecrest Project) may be regarded as a replication of the McDonald Training

Center Study (cf. above, p. 9ff.). As such, it confirms that several of the sub-variables in the Vocational Capacity Scale are satisfactory predictors. In the same way, one may assert that Elkin's second study (cf. above, p. 16) was a replication of the first and that on that account it shows that psychomotor and generally cognitive abilities can predict working capacity in both mildly and moderately mentally retarded young adults. Finally, one may recall that McCarron and Dial's study shows marked similarities with that of Wagner and Hawver (cf. above, p. 17). Though there is no reason to regard the former as being in a strict sense a replication of the latter, since neither the prediction nor the criterion variables were identical, one may nevertheless assert that together they confirm the possibilities of predicting the work performances of moderately retarded persons on the basis chiefly of psychomotor abilities.

## 5. The Author's Own Empirical Studies

### A. Summary of the initial position

In the immediately preceding introductory sections, my intention has been to indicate the initial position for the studies presented in the following pages. We have noted that modern society has taken on the responsibility for providing services to the mentally retarded. In series of follow-up studies, it has been possible to show that most of the individuals in this group have the ability to achieve and retain an acceptable degree of adjustment to work and to social life. However, they are more sensitive than other people to economic changes on the labour market, as well as to processes of change in commercial and industrial life and in production technique. In recent years, it has been observed more often than formerly that several mentally retarded persons who are unable to be fully self-supporting can nevertheless do productive work in a more or less sheltered environment. However, all of them are to a very large extent dependent on individual guidance.

In order to be able to give such guidance, it is necessary to know the prognostic values or the predictive efficiency of the critical variables. In spite of the numerous studies that have been made in this field, our knowledge of it is still to a large extent deficient. However, some advances have been made. Thus, it has been shown that, through a combination of different variables—chiefly in the psychomotor and generally cognitive sphere—it is possible to increase the accuracy of prediction. A more purposeful analysis of the criterion problems has also been commenced. However, we have as yet no successful cross-validations and the replications are few and far between. For this reason, further studies are required before we can assert that definitive results are available and can be applied in practice.

**B. Aims**

In view of this, I can now state that the main aim of my own empirical studies was to analyse the possibilities of predicting the work performances and the adjustment to work of mentally retarded young adults with varying potential for work—from those who are almost able to work independently as employees on the open market to those who barely possess the qualifications for taking part in productive work in a sheltered environment. This means that the methods applied in the studies must be adapted to both mildly retarded and moderately retarded persons.

Within this general framework, the present studies have the following aims:

(1) To analyse and determine the criterion variable, partly by observations and ratings of the working capacity and the adjustment to work and partly by objective measurements of performances in a number of tasks which are representative, as regards nature and degree of difficulty, of the ordinary tasks in sheltered workshops for the mentally retarded.

(2) To study which combinations of variables in the psychomotor and generally cognitive sphere yield the most accurate predictions.

(3) To study to what extent personality traits in mentally retarded persons—as these traits appear in systematic observations of behaviour at work—may help to predict the above-mentioned criterion.

(4) To carry on the studies in such forms that the procedures can, at least in part, be subjected to cross-validation.

**C. The concrete conditions for carrying out the studies**

The studies were carried out at the sheltered workshop for mentally retarded persons which the ALA Foundation ran in Uppsala from 1966 to 1970. This workshop came into existence with the explicit aim of serving as a pattern in various respects for other institutions. The manager and the supervisors attempted to carry on productive work in rational forms in this workshop but were not content with that alone. They regarded it as their principal task "to develop in each individual the maximum and most comprehensive ability, as regards performance and function" (Palmér 1968, (*a*) p. 35). From this point of view, it was of great importance that the majority of the employees lived in a home for boarders, where they received an all-round and purposeful social training as a complement to the work training.

The workshop was also intended to make technical and educational development work possible and to facilitate interdisciplinary research in various ways. Both the workshop manager and the supervisors were in complete agreement with these aims and played an important part in the research environment which was gradually established through daily contacts between the different parties. In consultations between the research director and the workshop manager, frequent

conferences were arranged concerning, *inter alia,* the methods of observation, the giving of instructions, and checks and evaluations in connection with the criterion measurements, which went on continuously for two whole years. Without this close collaboration, it would not have been possible to give the studies the comparatively wide scope which they had. Nor, in an environment less conducive to research, would it have been possible to carry out an objective study of work performances under such reatistic conditions as obtained in this case.

In one respect, however, the location of the studies at the ALA workshop was a restrictive factor. The workshop ceased its activities towards the end of 1970 and was converted into an "adjustment course" for the feeble-minded; it is now run by the National Board of Education. This was followed by such decisive changes, as regarded the composition of the staff and the design of the work, that we were no longer able to carry on the research project in the same forms as before. Otherwise, we had intended to complete the measurements of the predictor and criterion variables in a further sub-group of 16 individuals.

**D. The terminology employed**

The terminology used to indicate various degrees of mental retardation has varied from period to period and from country to country. In Sweden, as in several other countries, a rough demarcation line has long been drawn by distinguishing between the educable mentally retarded and the non-educable mentally retarded. The latter group have sometimes been alternatively described as the trainable mentally retarded and the dividing line between the two groups has, as a rule, been considered to be at an IQ of 50. Though this terminology is now in principle rejected by the majority of research workers, it is still used in a large number of accounts of empirical studies. It is also used in certain sections of the present account.

However, in parallel with it, I use the terminology which is recommended by the American Association on Mental Deficiency (cf. Grossman (ed.) *Manual on Terminology and Classification in Mental Retardation,* revised edition, 1973) and which involves a classification in four different degrees of retardation, namely, mildly retarded (IQ = 68–52), moderately retarded (IQ = 51–36), severely retarded (IQ = 35–20) and profoundly retarded (IQ below 20).

# Chapter II

# The Recruitment and Composition of the Experimental Groups

The study as a whole presupposed that I could have the experimental subjects at my disposal for comparatively long periods of time. The criterion measurements were made in the form of a testing programme, which was closely integrated with the regular production at the ALA Foundation's sheltered workshop. These two circumstances combined meant that the experimental subjects could only be chosen from among the employees at this workshop. In actual fact, "experimental subject" and "employee" became identical concepts, with certain exceptions which will be reported later. For this reason, it was of great importance that I was able, right from the beginning, to adopt a "research attitude" to the recruitment of the employees at the workshop and to make this whole question the subject of detailed discussions in the research advisory group mentioned earlier.

## 1. Etiological Classification as Principle of Selection

Certain voices were raised in the favour of the idea that, in the first place, I should endeavour to secure a composition of the worker group that would allow of comparisons being made between etiologically well-defined groups. A recruitment of this kind would undoubtedly also have been extremely interesting from the research point of view. Windle has given a clear justification for it, in pointing out the desirability, in general, of using homogeneous groups in prognostic studies: "The unique relations which exist for pure subgroups can only be established with certainty by studying these homogeneous subgroups" (p. 27).

However, this question is a complicated one, from several points of view. To begin with, it is by no means self-evident that differences in etiology imply differences in behaviour as a consequence. If we compare mongols with non-mongols, it certainly seems that this may be the case (Lindgren 1971, Ryan 1971), but numerous efforts to establish behavioural differences between endogenous and exogenous groups of mentally retarded persons have not succeeded (Windle, p. 135, Benton 1964, pp. 45 ff.). This latter fact may, of course, be connected with the generally recognized difficulties of making the ac-

tual classification by reliable methods. In all circumstances, such difficulties will become insurmountable obstacles when one is unable to observe the individual persons from birth but is reduced to using the medical and educational records which are available. As regarded the applicants for admission to the sheltered workshop, these records were detailed and comprehensive but nevertheless contained only to a limited extent data which could form the basis for an etiological classification.

In addition, there was a further circumstance which also made it impossible to consistently direct the recruitment of workers for the purpose of obtaining etiologically well-defined groups. The workshop was expected to serve a social purpose as a habilitation institute for the mentally retarded, chiefly in Uppsala county. Amongst other things, important financial conditions were dependent on this expectation being fulfilled. It was therefore necessary, in the first place, to allow the individual needs of habilitation measures to be decisive, as regarded admission. However, in order to obtain, within this framework, as representative a selection as possible, it was decided to carry out a systematic inventory of the needs in the county. A separate account of this inventory was given at the XIII Nordic Congress on Mental Retardation, held at Stockholm in 1967, but has only been published in summary form (Fränkel, Hedsjö and Palmér 1967).

## 2. Inventory of Needs and Criteria for Selection

As the foregoing remarks have shown, our purpose in making this inventory was chiefly to ensure that all the mentally retarded persons in the area who needed to be placed in sheltered work and who could be expected to profit from it in a meaningful way should also have the chance of being considered for employment in the workshop. However, the principle of need is very elastic and we therefore found it necessary to specify its meaning to a certain extent and, in addition, to formulate certain future criteria for selection, which would serve the purpose of making the group reeasonably homogeneous, as regarded age, intellectual capacity and previous schooling.

In close connection with these considerations, we also had to decide whether we should limit the recruitment to the "educable mentally retarded" or open the workshop also to the "non-educable mentally retarded". Several lines of argument were discussed in this connection. From the research point of view, it was possible to justify such a limitation by reference to the desirability of studying an intellectually homogeneous group. In practice, we could expect to probably avoid certain types of awkward problems by this means. The principal argument against the limitation was the knowledge that the "non-educable mentally retarded" group contains a fair number of individuals who after training can justify being given a place in a sheltered workshop and in some cases can even

become self-supporting (cf. above, pp. 5–6). If this was the case, workers should fairly obviously be recruited also from that group, since our aim was to shed light on habilitation problems from all sides, as regarded the mentally retarded.

The last-mentioned argument was decisive and in our further work we allowed for two groups of recruits—an educable group and a trainable group. The following selection criteria were drawn up for these groups.

*The educable group*

1. Age 16–35 years.
2. Registered in the register of the Board for Provisions and Services to the Mentally Retarded (BPSMR) or discharged during the last 5 years.
3. Completed schooling in a special school for basic comprehension teaching in the prescribed manner (cf. Grunewald 1969, pp. 20 ff.).
4. Achieved at some time an IQ of at least 50 on the Terman-Merrill scale.
5. Is not taking part in education within the framework of the special school (vocational training or vocational instruction) or in any other vocational education.
6. Is not obviously socially maladjusted (not the subject of action by the Child Welfare Committee or the Temperance Committee during the last 12 months on account of improper behaviour).
7. As regards adjustment to work, is in such a situation that placement in sheltered employment seems justified. (The situation was investigated at an interview with the employer or other suitable person and by rating the individual in accordance with a special schedule.)

*The trainable group*

1. Age 16–35 years.
2. Registered in the BPSMR register or discharged during the last 5 years.
3. Not completed schooling in a special school for basic comprehension teaching in the prescribed manner.
4. Not so poorly adjusted socially that placement in the ALA Foundation's home for boarders is impossible. (The person concerned should be able to cope with tasks connected with his personal hygiene and daily toilet, to dress himself (except for the more difficult operations, such as doing up buttons at the back, knotting his tie, etc.) and to feed himself.)
5. As regards adjustment to work, is in such a situation that placement in sheltered employment seems justified. (The person concerned must have taken part, with some success, in work therapy or industrial therapy at an institution or have performed tasks at home, such as clearing tables, giving a helping hand, etc.)

If we consider these selection criteria more closely, we may note the following points. The common features of the two recruitment groups were the age limits and the fact that the individuals included were registered in the BPSMR register or had been discharged from it. The lower age limit was fairly obvious, as no one could have completed his schooling in the basic grade of the special school before the age of 16. However, in actual fact no employee younger than 18 years was admitted. The fixing of the upper age limit at 35 years was justified by the fact

that it was desired to study a fairly homogeneous group, in which all the members were still of an age at which they were capable of development. Even this limit was shifted upwards in a few isolated cases. The requirement of registration guaranteed that all those admitted had at some time in their lives been investigated and in that connection had been assessed by experts as mentally retarded.

Other common features of the two groups—although they were not identical—were subsidiary criteria intended to limit admission to such individuals as were in need of or could profit from the planned habilitation work. As regarded the educable group, these subsidiary criteria involved a delimitation upwards in relation to those who could already be regarded as being well adjusted to working life or who were being trained for an occupation. (As is clear from item 6, there was here also a delimitation in relation to those who were obviously socially maladjusted, a step which was taken in order not to make the start at the new workshop too demanding for untrained staff. However, only in one or two exceptional cases was admission refused on this account and this cannot be considered to have been of any real importance.) As regarded the non-educable group, the corresponding subsidiary criteria in items 4 and 5 involved instead a delimitation downwards in relation to individuals who showed clearly unsatisfactory social adjustment or who were too poorly qualified to take part in productive work.

The result of these delimitations upwards and downwards was that the two recruitment groups, taken together, came to include mentally retarded young adults who varied, as regarded working capacity, between those who were on the verge of being able to function independently as employees on the open market and those who could only doubtfully be assessed as being qualified to participate in productive work in a sheltered environment.

How we finally drew the boundary line between the two recruitment groups has been indicated in the third subsidiary criterion, according to which the mentally retarded person was assigned to the educable group if he had completed his schooling in the basic grade of the special school and thereby shown ability to profit from theoretical instruction up to a certain level. Some of the individuals to whom this applied had, at the beginning of their schooling, passed through one or more grades in normal classes or special classes. If they had not in this way shown ability to receive theoretical instruction, they were accordingly assigned to the non-educable group. The majority of the individuals to whom this applied had been tested in special schools but had been transferred after varying periods to a "practical class" or a training class. A few had not received any theoretical instruction at all. As an, in principle, less important parallel criterion, the requirement was made that the person concerned should at some time have achieved an IQ of at least 50 on the Terman-Merrill scale, in order to be assigned to the educable group.

## 3. The Recruiting Procedure and the Problem of Generalization

The work on the inventory was carried out by a welfare officer, who was able to consult continuously with the officials of the BPSMR. From the available case records, lists of individuals who satisfied the given criteria were compiled. At separate meetings which the research team had with officials of the BPSMR and members of the Habilitation Board respectively, all these individuals were then discussed, with the aim of checking whether there were special circumstances which argued against the proposed placements. From these discussions, it became clear, on the whole, that the inventory had yielded a picture of the clientele which was in good agreement with the experience and the assessments of the responsible officials. However, these officials found themselves in duty bound to oppose the proposed placement in a small number of cases, with reference to such circumstances (known to them) as marked emotional disturbances, pronounced physical weakness or peculiarities in the social situation. When, afterwards, the individuals about whose suitability there was agreement were finally asked, some of them refused the placement, in spite of energetic persuasion. In this way, a further drop-out arose and I am therefore compelled to record that the social reality was found to include so many facets that it was difficult to carry out a study of the whole population or to ensure a composition of the experimental groups that could certainly be considered to be representative of it.

Similar observations were made, as regarded the recruitment of employees from the adjoining northern part of Stockholm county. In the densely populated municipalities in this part of the country, there were several other sheltered workshops and it could only be a question of receiving a small proportion of the mentally retarded persons who lived here and who were in need of placement in sheltered work. It was therefore found to be scarcely meaningful to carry out an inventory similar to that in Uppsala county and, for various reasons, such an inventory could not be carried out, but the above-mentioned selection criteria were strictly applied at each admission. They were also applied in the recruitment of isolated cases from other counties.

From the three recruitment areas mentioned above, viz. Uppsala county, the northern part of Stockholm county, and other counties, 37, 36 and 7 employees respectively were received during the period for which the study lasted. It is obvious that the total group of employees recruited in this way cannot be said to be a sample which satisfies the usual statistical criteria for allowing conclusions to be drawn concerning a population. On the other hand, there is scarcely any reason to assume that the composition of the group was determined to a marked extent by any other systematically operating factors than the given selection criteria. It is, in fact, more likely that the multiplicity of determinants in the ad-

*Table 2.* Comparison between all the employees at 24 out of the altogether 29 sheltered workshops in Sweden and the employees of the ALA Foundation, as regards the sex distribution and the distribution between the educable group and the trainable group.

| Employee group | Men | | Women | | Educable | | Non-educable | |
|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % |
| All employees at 24 sheltered workshops | 432 | 58.4 | 308 | 41.6 | 435 | 61.3 | 275 | 38.7 |
| Employees of the ALA Foundation | 47 | 58.7 | 33 | 41.3 | 52 | 65 | 28 | 35 |

mission situations was so large that sometimes one and sometimes another was decisive.

A certain amount of support for this view is offered by a comparison which may be made between the present experimental group and the great majority of workers at the other sheltered workshops for mentally retarded persons which were set up in Sweden in 1967. As these workshops served as regional reception centres, we may assume that they satisfied fairly well the existing needs for habilitation and that thus they contained within themselves the bulk of the population which we wish to study.

In 1967, I sent out an extensive questionnaire to the 29 workshops which were then fully established and obtained complete answers from 24 of them (Palmér 1968*a*). Amongst other things, the workshop managers were requested to carry out a productivity rating, the result of which confirmed that in this respect the workers showed the same wide variation as characterized the present experimental group. As Table 2 shows, there is also a remarkably good agreement as regards both the distribution between the sexes and the distribution between the educable group and the trainable group. Thus, in important respects, the present experimental group seems to represent well the population which we are interested in.

Finally, in the light of the above discussion, I would formulate my view of the problem of generalization as follows. Though I cannot consider myself entitled to generalize from the above data without reservations, it should, however, be possible to make them into a basis for tentative conclusions as regards those mentally retarded adults who fall within the criterion limits mentioned and commented on above.

## 4. Sample Characteristics

Finally, I shall in this section describe the experimental groups in greater detail with regard to age, sex, IQ and distribution between the educable and the traina-

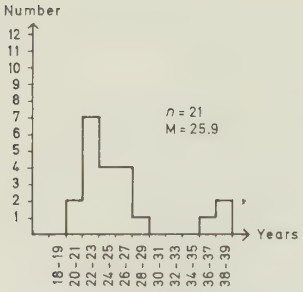



Fig. 1:1. Age distribution in men belonging to the trainable group.

Fig. 1:2. Age distribution in men in the total group.

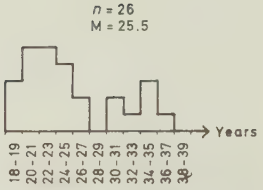



Fig. 1:3. Age distribution in men belonging to the educable group.

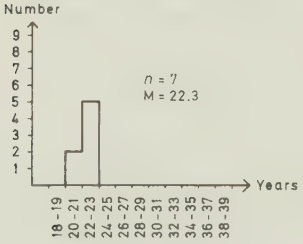



Fig. 2:1. Age distribution in women belonging to the trainable group.

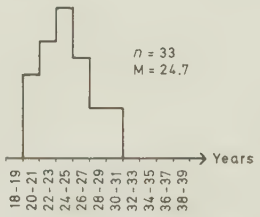



Fig. 2:2. Age distribution in women in the total group.

Fig. 2:3. Age distribution in women belonging to the educable group.

ble group. In addition, I shall also report the available etiological data.

The distribution by sex has already been given in Table 2, which shows that the group consists of 47 men and 33 women. Approximately this proportional distribution is not only found in workers in sheltered workshops in Sweden. As Wessman and Gustin's data show (p. 16), it is also characteristic of the total composition of the pupils in the special schools. These circumstances suggest that within this ability sphere there are certain general functional differences between men and women which would be worth a detailed analysis.

The age distributions will be found in Figs. 1 and 2 above, in the form of comparative histograms referring to the total, the educable and the trainable groups. However, in further studies, I shall, as a rule, have no reason to keep the two latter recruitment groups separate. On that account, let us devote our main attention to the total groups. It will be seen that the average age is nearly 25 years in both men and women. It may also be noted that the range of the age variable is somewhat greater in the men.

The IQ distributions in men and women are similarly shown in Figs. 3 and 4. I considered it natural and well justified to present these data, since impairment of

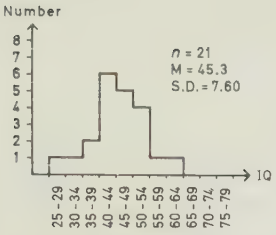



Fig. 3:1. IQ distribution in men belonging to the trainable group.

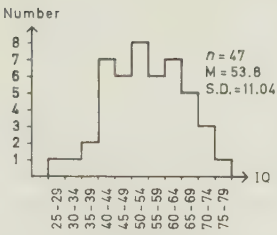



Fig. 3:2. IQ distribution in men in the total group.

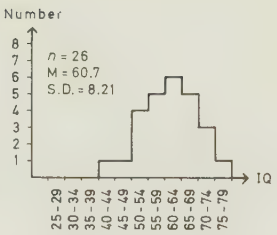



Fig. 3:3. IQ distribution in men belonging to the educable group.

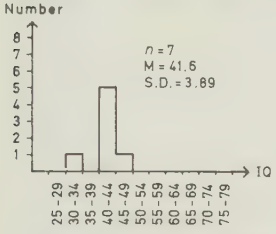



Fig. 4:1. IQ distribution in women belonging to the trainable group.

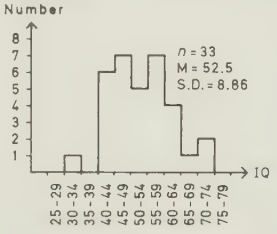



Fig. 4:2. IQ distribution in women in the total group.

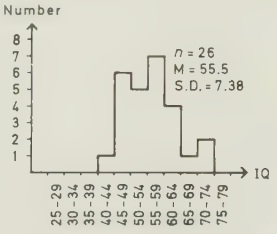



Fig. 4:3. IQ distribution in women belonging to the educable group.

intellectual function is generally regarded as the principal symptom of mental retardation. During the time when the experimental subjects were growing up, intelligence testing with the Swedish version of the Terman-Merrill scale was also being applied to a very large extent as the standard procedure in connection with placement in special schools and similar decisions. Also within the special schools, this method was frequently used for the purpose of following the pupils' development during their school-days. Thus, an examination of the case-record material for the present study showed that some experimental subjects had been tested by the Terman-Merrill scale as many as 10 times. However, the practice in this respect varied greatly and therefore it also proved impossible to compile test results from one and the same age for the whole group, as would otherwise have been desirable. When I tried the expedient of using the last test result for each individual, I found that this resulted in a marked dispersion of the ages from which the values were derived. For this reason, I considered it to be a better solution to calculate the average of the available test results in each individual case and to let this average, as the most reliable measurement, describe the intellectual level. For six of the experimental subjects, there were no Terman-Merrill values at all.

However, they had all been tested by either the Leiter, the Wisc, the CVB or the Grace Arthur scale, and in this context I considered that I could accept the results achieved in this connection as equivalent measurements of intellectual capacity.

Finally, I would point out that I am aware that the IQ values presented are not satisfactory measurements of the actual ability of the experimental subjects at the time when the study was carried out. For example, if I had wished to use them as predictor variables, I ought undoubtedly to have carried out fresh tests on all of them. However, as will become clear later, this was never my intention. I use these IQ values here only in order to give a general description of the experimental groups from the intellectual point of view, and for that purpose I consider them satisfactory.

As regards the mean values and standard deviations in the total groups, which I am also primarily interested in here, it is reasonable to compare them with Wessman and Gustin's data (1964, pp. 17 ff.). As these authors point out, their values are of particular interest, since they refer to four grades of the whole population of special-school pupils in Sweden. The comparison should chiefly refer to the pupils whom Wessman and Gustin designate "real special-school pupils", i.e. those who finish their period of compulsory school attendance in a special school. For these pupils, data and IQ results according to the Terman-Merrill scale are given both on admission to the school and on discharge from it. As far as the men are concerned, these average values amount to 57.2 and 56.2 respectively and, as far as the women are concerned, to 55.4 and 52.9 respectively. The corresponding standard deviations are 10.8 and 11.5 respectively in the male group and 10.4 and 12.2 respectively in the female group.

In view of this, the average values shown in Figs. 3 and 4 seem to be of the order of magnitude that might have been expected. It should be recalled that we excluded from the recruitment groups those who had already become adapted to working life and that we also admitted a few mentally retarded persons who had previously been found to be quite incapable of taking part in the instruction given in the special school. Both these circumstances together may be responsible for the fact that the experimental groups are, on the average, somewhat lower than the grade groups. The amount of the standard deviation in the group of men agrees very closely with the corresponding values in Wessman and Gustin's study. On the other hand, the range in the group of women is smaller.

Scrutiny of old records of employees, together with assessment of other data obtained in connection with medical examinations carried out at the place of work, revealed the cause of the mental retardation in only three cases out of 80. These three employees suffered from Down's syndrome (mongolism). One employee displayed Turner's syndrome and another a probable Elfin face syndrome. Nevertheless no decision was made to include these latter two in the

group in which the genesis of the mental retardation was known.

The remaining workers could be divided into two groups on the same grounds. To one of these were assigned 19 employees for whom no etiological data whatever were available. The other group included 56 workers whose case histories provided more or less detailed information on possible causes of the mental retardation. There was a very high frequency of mental retardation in parents or other relatives, but the form of the data suggested that this was only a matter of assumption rather than fully confirmed facts. Information on infectious diseases suffered during the first years of life was available in 12 cases, and data on perinatal injuries in five. However, these data are not such that they can be accepted as expressions of clearly proven causal relationships either. Against the background of this account, I could sum up by saying that the experimental subjects were characterised to an overwhelming degree by uncomplicated and etiologically unexplained forms of mental retardation.

Typical cases of double handicap were rare. However, the group included eight workers with more or less pronounced epilepsy. The same number suffered from noticeable stammering or other speech defects. There were five cases of impaired vision or hearing disability and two of organic heart defects. Slight mental disturbances were noted in five workers. Finally we recorded one case each of morbus Bechterew, scoliosis and bone shrinkage with spasticity.

In general, especially considering the number of days of absence from work the state of health may be described as good. Finally I should mention that the administration of drugs to the workers was greatly restricted.

# Chapter III

# Prediction Variables: Selection and Results of Measurements

From the summaries of previous prognostic studies mentioned in Chapter I, it is clear that attempts have been made to determine the prognostic value of a very large number of different factors. Windle classifies these factors and distinguishes four different groups, viz. (1) demographic characteristics, (2) patient abilities and disabilities, (3) institutional experiences, and (4) family and community factors (Windle, p. 42). We have already found that it has been possible in only a comparatively few cases to demonstrate reliable and unequivocal correlations between such factors and the criteria of successful vocational and social adjustment. The reasons for this finding are, of course, of many different kinds, but it nonetheless gives us cause to thoroughly examine the question of what variables—to use Wolfensberger's terminology—deserve to be called "hard predictors" and can consequently be adjudged to be worth further investigation.

However, the very multiplicity of the conceivable variables makes it necessary to impose a certain limitation on oneself from the beginning. In the present study, I worked from starting-points in differential psychology and asked myself what individual characteristics can predict working performances and adjustment to work in mentally retarded persons. Thus, without questioning their importance—which in reality is quite indisputable—I leave out of account here such factors as the applied methodology in habilitation work and the existing social relations in the home, at the workplace and during leisure hours.

## 1. Psychomotor, Spatial-Inductive and Numerical Abilities

### A. Choice of test instruments

The survey of old and new studies given in Chapter I argues that psychomotor and generally cognitive abilities are the individual variables which may be expected to have the greatest predictive efficiency. It is justifiable also from another angle to examine in further studies the possibilities of predicting working performances and adjustment to work on the basis chiefly of psychomotor abilities. In

the individual guidance of the mentally retarded, it is essential to use test methods, not chiefly in order to establish the magnitude of the functional impairments that may be present but in order to discover the positive points of connection and to trace out the potential resources. In this approach, it will be natural to study primarily the functional spheres in which these individuals achieve relatively the best performances. It has long been known that this is the case precisely with regard to psychomotor abilities.

Kebbon has recently been able to confirm this fact in studies under controlled conditions, in which he compared the performances in 16 tests of a group of mentally retarded adults ($n = 183$) and a group of adults of normal intelligence ($n = 92$). He summarizes the results as follows:

> The retarded adults examined in this study differed least from normal controls in performance on manual tests, their mean scores on these tests lying about once the standard deviation under the normal mean. They did worse on the speed tests and still worse on the spatial-inductive tests, where their means were about twice the standard deviation lower than the normal means in both cases. They did worst of all in comparison with normal subjects in scores on verbal and numerical tests, particularly on multiplication (Kebbon 1965, p. 71).

Thus, I had good reasons for primarily using psychomotor abilities as prediction variables. From the beginning, it was also my intention to limit myself entirely to these abilities in this study. However, there are a good many reasons why in a battery of predictors one should also include spatial-inductive abilities and certain related skills in the cognitive sphere. Thus, the above quotation from Kebbon indicates that these abilities also display a comparatively good developability in mentally retarded persons. In this respect, they occupy an intermediate position between psychomotor abilities, on the one hand, and verbal abilities, on the other. It is also commonly assumed in industrial psychology that there are positive correlations between these and working capacity. Kebbon presents data which confirm that this is the case in mentally retarded workers. Thus, he calculates the correlations between factor scores and ratings of working capacity in two groups of mentally retarded persons and reports them in a table (Kebbon 1965, p. 30). If we examine this table, we find that the working capacity is most closely correlated with the calculated factor scores in the psychomotor and the numerical spheres (.43 and .51, and .47 and .57 respectively), then with the factor scores in the spatial-inductive sphere (.39 and .49) and least closely with the factor scores in the verbal sphere (.27 and .47).

These results indicate in general that there are comparatively good possibilities of predicting working capacity in mentally retarded adults on the basis of ability measurements. They also argue that, in the selection of predictors, we should not limit ourselves to psychomotor abilities but should also include numerical and spatial-inductive abilities. Properly speaking, it might also be justifiable to include

verbal abilities in a multidimensional predictor battery, but, as a rule, this possibility has to be excluded, owing to the difficulties of securing sufficiently differentiating measuring methods.

In the final selection of instruments for the ability measurements, I was able to benefit from Kebbon's previously quoted factor-analytical studies, which resulted in the design of the "Ulleråker Battery" (UB) (Kebbon 1968). From this, I derived the sub-tests which had proved to measure the psychomotor, spatial-inductive and numerical factors. However, I omitted at an early stage the spatial-inductive "Wire Design Copying" and the numerical multiplication test, which proved—not unexpectedly—to be too difficult for our experimental groups and consequently to give rise to extremely skewed distributions.

In order to accomplish my primary purpose of covering the psychomotor functional sphere more comprehensively, I added to the test battery some tests which had been developed fairly recently at the Psychotechnical Institute in Stockholm, namely, "Wire Labyrinth", "Threading" and "Plates" (cf. the manuals published by Skandinaviska Testförlaget 1967). For the same purpose, I also included the "Motor Test of Tracing Ability" (MTTA), "Tapping", which is a sub-test included in Trankell's handedness scale (Trankell 1960), and "Copying", which is included in Egidius's Group Test for Grade 1.

In order to supplement the Ulleråker Battery in the spatial-inductive sphere, I also included "Concepts", which is a sub-test in a knowledge test for special-school pupils prepared under the direction of Witting (Domme and Lindahl 1966), Raven's Progressive Matrices (the coloured version), the Leiter International Performance Scale and "Figure Classification" (Dureman and Sälde 1959). As a further supplement, I included in the numerical sphere "Addition (Witting)" and "Subtraction (Witting)", which are both sub-tests in the above-mentioned knowledge test for special-school pupils, and in the linguistic sphere "Vocabulary", which, like "Copying", is included in Egidius's Group Test for Grade 1. Finally, a simple test of reaction time, described by Elfström and Granberg (1968), was administered.

**B. Test administration and evaluation of performances**

The recruitment of employees for the ALA Foundation workshop commenced in February 1967 and then proceeded in stages until the full number (60—70 employees) was achieved in the autumn of 1968. The ability measurements were carried out successively in mainly the same order for all participants. I acted as the test director and was aided by a number of assistants trained in psychology, who had all for varying periods been engaged in various aspects of the habilitation work. All the measurements took place in close connection with the working premises and in ordinary working time. I consider that these circumstances are

important, as I gradually became more and more convinced that for mentally retarded individuals it is essential that they should be familiar with the external environment in which ability measurements take place and that they should recognize in the psychologist a person whom they have met and got to know in other connections.

In the administration of the various sub-tests, the instructions and recommendations in the published manuals were followed in principle. However, in a couple of cases minor modifications of the procedure were made, in order to adapt it better to the experimental group. This was the case in the "Threading" test.

In "Threading", 150 discs have to be threaded as quickly as possible on two vertical rods of different thicknesses. The discs differ with respect to diameter and colour, in that half of them fit on one rod and half on the other. In this way, the test also includes a simple sorting task. I omitted this sorting test and allowed the experimental subjects to first thread the large discs on one rod and then the small discs on the other, in order to make the test more refined from the psychomotor point of view. The instructions were, of course, adapted to the altered procedure, clarified in one respect and given the following wording:

*Large discs.* You are to put these discs on this rod [pointing to the rod] as quickly as you can. You must take only one disc at a time from the bowl and then put it on the rod. In other words, you must put on all the discs in the shortest possible time. Is there anything you want to ask about? Begin now [timekeeping was commenced].

*Small discs.* Now you are to put on these small discs instead. They are to be put on this rod [pointing to the rod]. Remember that you may pick up only one disc at a time. Begin now [timekeeping was commenced].

The raw scores were also in principle added together in the way prescribed in the manuals. However, in this connection also certain changes should be mentioned. Thus, in the two handedness tests, MTTA and "Tapping", the scores for the performances with the right and the left hand were added together to form a total.

It is more important to point out that in certain tests I was not content to calculate a total score in the usual way and to let this represent a general measurement of skill; in addition, I also calculated the intra-subject variability of the performances. The use of such measures of variability as prediction variables may be justified on the basis of the following considerations. The variability is an expression of the degree of consistency in the behaviour. It seems probable that this characteristic is of importance in concrete working situations. Baumeister and Kellas (1968) suggest this, after reporting studies which yielded clear evidence that the variability as regards reaction times—both between and within individuals—is greater in mentally retarded persons than in persons of normal intelligence: "These subjects are unable to meet environmentally imposed demands in a consistent manner".

There proved to be reasonable chances of calculating intra-subject variability

in three of the tests, namely, the simple test of reaction time, "Wire labyrinth" and "Tapping".

### C. Estimation of reliability

Cronbach (1969), amongst others, has pointed out that the reliability may be different at different levels of performance and in different groups of individuals. Thus, although a test method has proved to satisfy reasonable requirements in groups of people of normal intelligence, it is desirable to test its reliability afresh, if one wishes to apply it to mentally retarded persons. Kebbon (1965, 1968) has taken this observation to heart and published reliability coefficients for most of the sub-tests in the Ulleråker Battery, as well as for Raven's Matrices and "Figure Classification", calculated both on groups of people of normal intelligence and on groups of mentally retarded persons. He observes that "all the coefficients are around 0.80 or higher for both the normal group and the retarded group, except for 'Memory for design' (0.75 and 0.57 respectively)".

For the other tests which I used in this study, I have—with one exception—calculated reliability coefficients and brought them together in Table 3. The exception is the Leiter International Performance Scale, for which, in collecting the data, I neglected to take the necessary steps to determine the reliability. As Table 3 shows, in most cases the calculations were carried out by the split-half method. As regards the detailed procedure, the following remarks may be made.

In "Wire labyrinth", I, like the author of the manual, let the first subsidiary operation serve as an exercise and in this way obtained two equally long series of seven trials each. I applied the same procedure in "Copying". In "Threading", I correlated the time taken to thread the large discs with the corresponding time taken to thread the small discs, and in "Plates" the reliability was determined as the correlation between the measurements in the two performances of the task, as the manual prescribes. In the MTTA, I followed the same procedure as the test designers and calculated separate coefficients for the right and the left hand. In addition, however, I also correlated the performances with each hand and regarded the coefficient obtained in this way as an expression of the reliability of the total value used in the study. The same procedure was followed in "Tapping", apart from the fact that I did not carry out separate calculations for each hand. The score in the simple test of reaction time is the average of 200 determinations. In order to calculate the reliability of this value, I divided the total number of determinations into 20 series, each of 10 experiments. Every other series was then assigned to the respective halves within which the mean values were calculated and correlated. "Concepts", "Addition (Witting)" and "Subtraction (Witting)" were all administered without applying any time limit. This made it possible to determine the reliability in these cases with the aid of the Kuder-Richardson formula KR20. Finally, as regards the measurements of variability,

*Table 3.* Reliability coefficients in a number of prediction variables. MTTA = Motor Test of Tracing Ability. SH = Split-half method corrected by the Spearman-Brown formula. KR = Kuder Richardson.

| Variable | Coefficient | Method of calculation |
|---|---|---|
| Wire labyrinth | .97 | SH |
| Threading | .95 | SH |
| Plates | .94 | SH |
| MTTA, right hand | .98 | SH |
| MTTA, left hand | .98 | SH |
| MTTA, right and left hands | .96 | SH |
| Tapping, right and left hands | .97 | SH |
| Copying | .93 | SH |
| Reaction time | .96 | SH |
| Concepts | .79 | KR |
| Addition, W | .94 | KR |
| Subtraction, W | .97 | KR |
| Vocabulary | .73 | SH |
| Reaction time, S.D. | .94 | SH |
| Wire labyrinth, S.D. | .72 | SH |
| Tapping, S.D. | .87 | SH |

the same procedure was applied in principle as in the corresponding measurements of performance.

A scrutiny of Table 3 will show that "Vocabulary" and "Concepts" yield the comparatively low values of 0.73 and 0.79 respectively. Evidently, in their present forms, these two tests are not particularly suitable for use among the mentally retarded. Furthermore, it may be noted that one of the measurements of variability, namely, "Wire labyrinth, S.D.", shows the unsatisfactorily low value of .72. This would seem to be connected with the fact that the measurements is based on a numerically too limited series of observations. In other respects, however, the coefficients are quite satisfactory and in most cases remarkably high. They consequently show—like the corresponding values in Kebbon's work—that there are good chances of obtaining reliable results in ability measurements among mentally retarded adults.

### D. Average performance level

The considerations reported in the foregoing pages, as regarded the selection of test instruments and the evaluation of performances, have enabled me to report results for all the individuals in the experimental groups in the following variables, which are reproduced here under the appellations and designations which will be consistently used in this account.

(*a*) *Psychomotor abilities*

1. Crawford screws
2. Crawford pins
3. O'Connor Test
4. Mark-making
5. Wire labyrinth
6. Threading
7. Plates
8. Motor test of tracing ability (MTTA)
9. Tapping
10. Copying

(*b*) *Reaction time*

11. Simple test of reaction time

(*c*) *Spatial-Inductive abilities*

12. Memory for design
13. Surface development
14. Block design
15. Raven's Progressive Matrices (coloured version)
16. The Leiter International Performance Scale
17. Figure classification
18. Concepts

(*d*) *Numerical abilities*

19. Addition (Ulleråker Battery (UB))
20. Addition (Witting (W))
21. Subtraction (Witting (W))
22. Highest number
23. Identical forms

(*e*) *Linguistic skills*

24. Vocabulary

(*f*) *Measures of intra-subject variability*

25. Reaction time, S.D.
26. Wire labyrinth, S.D.
27. Tapping, S.D.

The median values, mean values, standard deviations and ranges of variation in men and women in these variables will be found in Table 4. The most con-

*Table 4.* Results of ability measurements, expressed as medians, means, standard deviations and ranges.

| Variable | Men ($n$ = 47) | | | | Women ($n$ = 33) | | | | Sex-diff. $t$-value |
|---|---|---|---|---|---|---|---|---|---|
| | Mdn | Mean | S.D. | Range | Mdn | Mean | S.D | Range | |
| 1. Crawford screws | 12.60 | 12.40 | 5.32 | 25 | 7.14 | 8.73 | 4.47 | 17 | 3.20 |
| 2. Crawford pins | 11.86 | 11.32 | 6.11 | 26 | 8.67 | 9.73 | 5.41 | 21 | 1.19 |
| 3. O'Connor test | 29.50 | 32.15 | 15.38 | 58 | 38.75 | 36.24 | 11.50 | 45 | −1.28 |
| 4. Mark-making | 79.83 | 74.19 | 33.66 | 205 | 96.90 | 94.06 | 30.72 | 150 | −2.66 |
| 5. Wire labyrinth | 870.50 | 1189.79 | 890.60 | 4545 | 988.00 | 1253.85 | 674.12 | 3205 | 0.34 |
| 6. Threading | 405.50 | 434.51 | 139.61 | 695 | 373.36 | 410.76 | 106.14 | 579 | −0.81 |
| 7. Plates | 13.13 | 15.28 | 9.61 | 37 | 13.67 | 14.27 | 8.07 | 31 | 0.48 |
| 8. MTTA | 675.50 | 865.26 | 797.67 | 4108 | 908.83 | 915.00 | 590.16 | 2853 | 0.30 |
| 9. Tapping | 189.50 | 237.55 | 143.21 | 74.9 | 185.50 | 219.61 | 103.03 | 35.5 | −0.61 |
| 10. Copying | 7.67 | 7.68 | 4.49 | 15 | 5.25 | 5.82 | 3.61 | 13 | 2.00 |
| 11. Reaction time | 61.10 | 69.34 | 38.95 | 237 | 62.75 | 65.67 | 19.79 | 77 | −0.49 |
| 12. Memory for design | 4.45 | 10.85 | 20.82 | 110 | 0.25 | 6.36 | 8.28 | 25 | 1.16 |
| 13. Surface development | 4.42 | 6.47 | 5.82 | 23 | 4.80 | 5.12 | 3.14 | 16 | 1.20 |
| 14. Block design | 6.75 | 7.89 | 5.84 | 21 | 3.35 | 5.97 | 5.45 | 19 | 1.47 |
| 15. Raven's Progressive Matrices (coloured version) | 20.33 | 20.13 | 6.03 | 24 | 15.75 | 16.09 | 4.78 | 20 | 3.26 |
| 16. The Leiter International Scale | 76.75 | 80.13 | 19.63 | 76 | 77.75 | 79.12 | 15.04 | 79 | 0.24 |
| 17. Figure classification | 6.33 | 7.30 | 6.51 | 20 | 6.25 | 6.39 | 5.68 | 20 | 0.63 |
| 18. Concepts | 9.65 | 8.83 | 2.50 | 10 | 10.00 | 9.00 | 2.44 | 8 | −0.30 |
| 19. Addition, UB | 2.67 | 6.17 | 7.13 | 27 | 6.00 | 7.33 | 8.09 | 26 | −0.67 |
| 20. Addition, W | 18.25 | 19.00 | 14.89 | 40 | 28.75 | 24.61 | 12.59 | 40 | −1.74 |
| 21. Subtraction, W | 6.00 | 12.15 | 11.90 | 33 | 11.74 | 13.24 | 10.28 | 33 | −0.42 |
| 22. Highest number | 10.25 | 15.43 | 13.92 | 62 | 13.13 | 15.09 | 11.56 | 46 | 0.10 |
| 23. Identical forms | 11.13 | 12.00 | 8.93 | 30 | 10.00 | 9.61 | 6.90 | 25 | 1.28 |
| 24. Vocabulary | 13.54 | 12.94 | 2.37 | 12 | 13.78 | 13.15 | 2.22 | 10 | −1.62 |
| 25. Reaction time, S.D. | 13.50 | 16.89 | 14.19 | 71 | 10.75 | 15.09 | 10.47 | 38 | −0.61 |
| 26. Wire labyrinth, S.D. | 17.36 | 24.64 | 22.42 | 115 | 25.50 | 39.85 | 33.99 | 153 | 2.38 |
| 27. Tapping, S.D. | 22.50 | 29.98 | 28.14 | 16.6 | 17.50 | 25.67 | 20.90 | 7.6 | −0.74 |

*Table 5.* Mean values and stanine scores in 10 test variables in a normal group of men and in two groups of mentally retarded men at different levels.

| Variable | Normal Ia (acc. Kebbon) | | Retarded Ib + II (acc. Kebbon) | | Own exptl group $n = 47$ | |
|---|---|---|---|---|---|---|
| | Mean | Stanine score | Mean | Stanine score | Mean | Stanine score |
| Crawford screws | 22.20 | 7 | 15.80 | 5 | 12.40 | 4 |
| Crawford pins | 25.46 | 8 | 17.76 | 5 | 11.31 | 3 |
| O'Connor test | 61.82 | 8 | 44.75 | 5 | 32.15 | 3 |
| Mark-making | 178.88 | 8 | 114.04 | 5 | 74.19 | 3 |
| Memory for design | 9.39 | 9 | 2.50 | 5 | 1.04 | 4 |
| Surface development | 23.08 | 9 | 9.27 | 5 | 6.47 | 4 |
| Block design | 27.28 | 9 | 11.77 | 5 | 7.89 | 4 |
| Addition, UB | 39.59 | 8 | 17.89 | 5 | 6.17 | 3 |
| Highest number | 65.08 | 8 | 36.79 | 5 | 15.43 | 3 |
| Identical forms | 33.77 | 9 | 16.88 | 5 | 12.00 | 4 |

spicuous feature is, of course, the consistently low average level of performance. In order to illustrate the nature and magnitude of this lowness of level, I have brought together in Table 5 the mean values in the male group in the ten test variables taken from the Ulleråker Battery (UB) and the corresponding values which Kebbon has reported for a group of men of normal intelligence, on the one hand, and a group of mentally retarded men with an average IQ of 70, on the other (Kebbon 1965, p. 101). Table 5 also gives the stanine scores, which are derived from the standardization of the results in the group of mentally retarded persons studied by Kebbon (cf. Kebbon 1968, p. 35). If, in comparing these values, it is remembered that my experimental group has an average IQ of 54, the obvious difference between the two groups of mentally retarded persons will be both expected and fully understandable.

**E. Ability to differentiate and forms of distribution**

In spite of the low average level of performance, the frequency distributions drawn up[1] show that most of the tests used possess a fairly satisfactory ability to differentiate. This observation applies particularly to the tests of psychomotor ability. However, as was expected, a certain tendency to skewed distributions may be discerned even in them. We also find in the psychomotor tests in which time is the dependent variable—"Wire labyrinth", "Threading", MTTA and "Tapping"—the characteristic circumstance (cf. Edwards 1968, pp. 109 ff.) that individual experimental subjects were extremely slow, with the result that comparatively large gaps appear towards the negative end of the distribution. This is also the case in the simple test of reaction time.

[1] They are available in duplicated form and may be obtained from the author on request.

In the spatial-inductive ability tests, the ability to differentiate is greatest in the Leiter International Performance Scale, Raven's Progressive Matrices, "Surface Development" and the "Block Design". In "Memory for Design" and "Figure Classification", the tendency to positively skewed distributions appears more strongly, while "Concepts" and "Vocabulary"—which were both designed for individuals whose performances were weak—were obviously too easy and therefore gave rise to negatively skewed distributions.

As was expected, comparatively large deficiencies in ability to differentiate appear in the tests of numerical ability. This is most obvious in "Addition (UB)". "Addition (W)" and "Subtraction (W)", which were designed for use among the mentally retarded, differentiate better but instead display a certain tendency to a U-shaped distribution. "Highest Number" and "Identical Forms" differentiate acceptably, though the tendency to positive skewness is clearly evident.

Finally, as concerns the three measurements of intra-subject variability, it may be said that, on the whole, they have proved to possess the same ability to differentiate as the corresponding average or total scores in the same test. This means that a certain tendency to skewed distributions is to be found also in them.

After examining the distribution forms obtained, I found it best to analyse the material without taking any steps to transform or to standardize the distributions. The experimental subjects in this study are a highly selected group and the reported tendencies to skewness are consequently expected and quite understandable. However, in order not to be compelled to correlate quite unequal distributions with each other, I took the precaution of turning the negatively skewed distributions in "Concepts" and "Vocabulary" so that they also took on the positively skewed form.

Finally, I would also point out in this connection that the above-mentioned deficiencies in ability to differentiate in these tests reflect the current situation in the test methodology of differential psychology. There is a great lack of satisfactory test methods which, as regards degree of difficulty and functioning in other respects, are adapted to the mentally retarded. In my endeavour to tackle the problems of prediction, I was not able to do very much to remove this defect. I had to be content with selecting from among the available test methods those which appeared to be least objectionable from the above points of view and with making such minor modifications as I have described above. In the end, I had to tolerate certain deviations from the desirable normal distributions.

**F. Sex differences and intercorrelation patterns**

If we study Table 4 with regard to the possible sex differences, we find, on initial examination, that men and women show practically the same average values in several of the 27 variables. In some of these variables, however, fairly obvious sex differences seem to appear. It was of interest to study this more closely, since

I had previously been able to show that the groups of men and women were recruited in the same way and were markedly alike, as regarded age and general intelligence level. I therefore carried out a multi-variate analysis of variance, which included all 27 variables. This resulted in a $F$ value of 2.98, which means that there is a probability of less than 1 in 100 that the observed differences are throughout a matter of chance. I therefore had reason to go further and pinpoint these differences by calculating their respective $t$-values, which are given in the last column of Table 4. Their magnitude confirms that in most cases the differences are negligible. However, the men show significantly better results in the following four variables:

| | |
|---|---|
| Raven's Progressive Matrices | $P = < 1/00$ |
| Crawford screws | $P = < 1/00$ |
| Wire labyrinth, S.D. | $P = < 5/00$ |
| Copying | $P = < 5/00$ |

On the other hand, the women perform significantly better in "Mark-making" ($P = < 1/00$). The differences in "Addition (W)" and "Vocabulary" also go in the same direction, though they do not achieve full statistical significance.

It is worth mentioning that these results point in the same direction as the corresponding results in Aldrin and Hallerstedt's work (1966). Although neither their work nor the present work can be made the basis of general and definitive conclusions, they nevertheless provide comparatively strong proof that mentally retarded men produce better performances in certain psychomotor skills of the kind represented by "Crawford Screws" and in certain spatial functions. Women, on the other hand, seem to achieve better results in such psychomotor skills as are involved in "Mark-making" and the O'Connor Test. They also show a tendency to be superior in numerical functions.

The product-moment correlations between the ability variables will be found in Appendices A:1-2 on p. 178–179. If we examine these tables, we shall fairly soon find that in many cases the coefficients are higher in men than in women. However, this tendency is not equally strong in all variables. I have tried to make this clear in Fig. 5:1–4. A study of these diagrams will show the following features. In the simple test of reaction time, "Tapping" and "Vocabulary", the coefficients are approximately equal in men and women and are also characterized by a considerable dispersion in both groups. In the O'Connor Test, the Leiter Scale and "Mark-making", we may note a less degree of dispersion, particularly in the men, and here the tendency to higher coefficients in the male group appears with marked strength. The two other groups of variables occupy an intermediate position from this point of view. There seems to be scarcely any clear correlation between the above-mentioned tendency to higher inter-

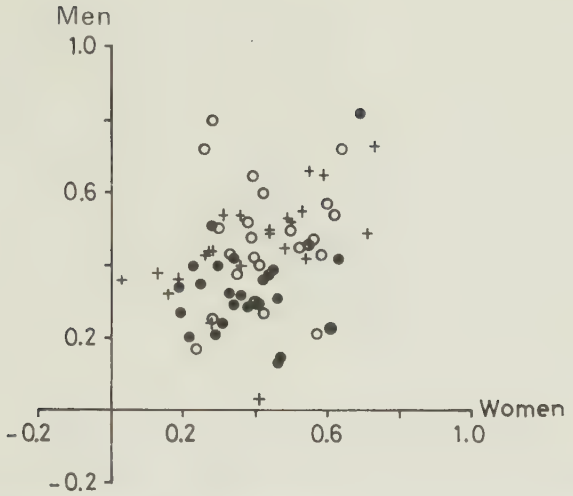



Fig. 5:1. Intercorrelations in men and women among all other prediction variables and Reaction time (●), Tapping (○), and Vocabulary (+).

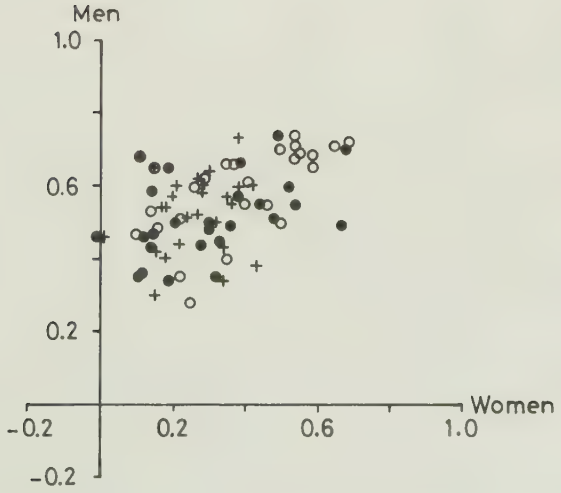



Fig. 5:2. Intercorrelations in men and women among all other prediction variables and O'Conner test (●), the Leiter Scale (○) and Mark-making (+).

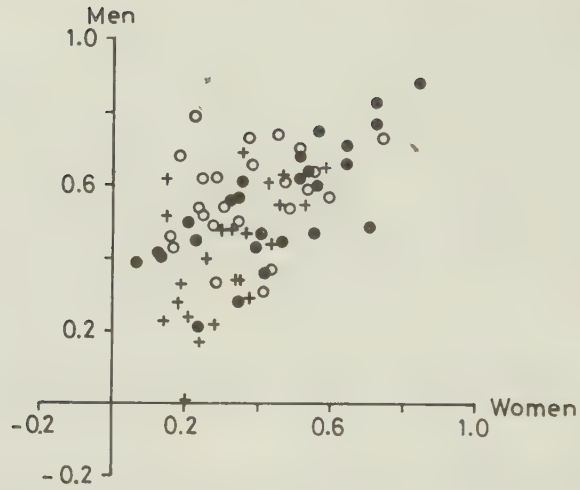



Fig. 5:3. Intercorrelations in men and women among all other prediction variables and Addition, W (●), MTTA (○) and Surface development (+).

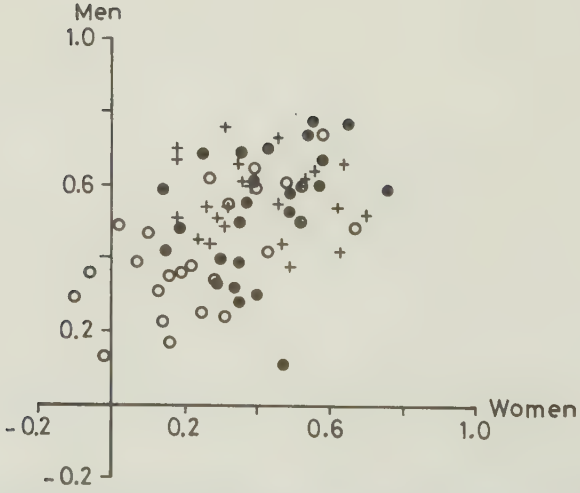



Fig. 5:4. Intercorrleations in men and women among all other prediction variables and Block design (●), Threading (○) and Plates (+).

correlations in the men and the factor affiliation of the test variables, although there are signs that this tendency appears most strongly in certain psychomotor and spatial variables.

The differences demonstrated between men and women in the intercorrelation pattern of the ability variables may, at least to a certain extent, be a result of the less degree of variability in the female group. However, I considered that they justified my treating men and women as two different groups in the following data analyses—as long as these analyses referred to the whole experimental group. I shall return to this question later on, in analysing the correlations between prediction variables and criterion variables.

## 2. Personality Characteristics

### A. Measuring problems

If we wish to investigate the possibilities of predicting the work performances and the adjustment to work of mentally retarded persons on the basis of individual variables, we should not be justified in restricting ourselves to abilities of various kinds. Experience of daily life convinces most people who have reason to reflect on the matter that personality factors are also of importance in this connection. Another circumstance which argues in favour of this view is that psychologists have generally noted that performances are regularly conditioned both by cognitive ability and by motivation (Holm and Palmér 1972). If we go back once more to the works of Windle, Patterson and Wolfensberger for enlightenment on this point, we find that all these three authors ascribe a determinative predictive efficiency to personality factors and emphasize the importance of further analysing this problem complex in fresh studies. Patterson (p. 142) formulates this view as follows: "If personality is, as seems to be the case, the crucial factor in adjustment and success of the mentally retarded, then it would appear to be desirable to use measures of personality in evaluation and prediction".

However, all three authors are also well aware that there are particular difficulties here, as regards measuring technique, since most personality tests make great demands on reading ability and self-reporting skills. I myself have been able to verify, in connection with my practical habilitation work, that these tests can consequently be used only to a very limited extent in studies of the mentally retarded. I have tested, without prejudice, such well-known instruments as the MMPI, Rorschach, TAT, Holzman Inkblot Technique and different versions of sentence-completion tests, only to discover that none of these methods is generally applicable.

Of course, this view does not deny the importance of making further attempts to adapt particularly the projective methods for use among the mentally retarded. As Cobb (p. 144) points out, these possibilities increase if one limits one's in-

vestigations to mild forms of retardation. In view of the fact that both Cobb and Patterson speak comparatively optimistically about the possibilities of using a sentence-completion test as a prediction instrument among the mentally retarded, it is worth mentioning here that I actually obtained the most promising results with a Swedish modification of Rotter's Sentence-completion Test. I therefore consider that a test of this character may well, after a certain amount of revision, be able to defend its place as an aid in psychodiagnostic work among the mentally retarded. In the present connection, however, I encountered too great difficulties in my endeavours to evaluate the responses in a reliable manner and therefore found myself compelled to give up the idea of including it among the prediction variables.

Accordingly, when I found it impossible to use conventional personality tests for these purposes, I decided instead to start from systematic observations of the behaviour in the working environment. In a certain sense, this may be regarded as a makeshift, but I consider that I can now state that in actual fact the procedure involves an easily defensible alternative, which may also be justified from the theoretical point of view.

In his critical analysis of the "trait-state approach" in personality psychology, which is the theoretical frame of reference for conventional personality tests, Mischel (1968) asserts that "large-scale efforts to predict behavior from personality inferences have been strikingly and consistently unsuccessful". The reason for this is not measuring errors or other shortcomings in carrying out the investigations—though such shortcomings naturally existed—but the circumstance that behaviour is really characterized by a high degree of specificity and a low degree of consistency. This has been known ever since Hartshorne and May presented their well-known *Studies in the Nature of Character* during the period 1928—30, but researchers have been unwilling to draw the natural conclusions from it and abandon the initial theoretical model. Mischel thinks that it is necessary that they should do so and also asserts that the prediction problem can be tackled with greater success from "social-behavior theory", which does not seek the determinants of behaviour in "stable response predispositions" but in "the conditions that covary with the occurrence, maintenance and change of behavior". According to this theory, behaviour tends to be "extremely variable and unstable, except when stimulus conditions and response-reinforcement relations are highly similar and consistent". In good agreement with this view are also the findings in several investigations which show that it is precisely observations and measurements of relevant previous behaviour which can most successfully predict future behaviour in similar situations.

Even in the research which has been specially devoted to examining the possibilities of prediction, as regards vocational and social adjustment among the mentally retarded, we may find support for the endeavour to design prediction

variables from direct observations of behaviour in the working environment. Thus, Patterson has made quite definite pronouncements in that direction. One of his main recommendations includes a "concentration of efforts on observing and evaluating relevant psychological and personality factors in a working environment or a simulated work situation." Wolfensberger also takes the same line when he argues that research workers should use "social, background and observational data" in various ways.

**B. Analysis of sub-scales in a general observation schedule of check-list type**

As a form for use in observing the working behaviour, I first chose to design a check list, which finally came to consist of 155 items. I asked all the supervisors at the workshop to assist by mentioning examples of behaviour of both a positive and a negative kind and with a concrete and specific meaning, on the basis of their practical experience. I obtained further suggestions for statements of this character by my own preparatory observations and by making an inventory of previously published check lists. In these different ways, I tried to discover and formulate a sufficiently large number of statements referring to three spheres of behaviour, to which I gave the following designations in the initial situation: (1) interest, persistence and motivation to work, (2) manner of carrying out working operations and orderliness in the workplace, and (3) emotional and social adaptation in the working environment.

In each sphere of behaviour, I tried to obtain a fairly even distribution of statements with positive and negative meanings. I also endeavoured to formulate the statements in such a way as to represent different degrees of emphasis on the behaviour in question. The final selection of items and the definitive formulations were not decided on until two preliminary versions had been practically tested on a small scale.

The 155 statements were compiled to form a "general observation schedule at a sheltered workshop for mentally retarded persons". In order to counteract the formation of response sets, I saw to it that items referring to different spheres of behaviour and items with positive and negative formulations were mixed. Finally, the form was prefaced by the following written instructions:

> The schedule consists of a number of assertions or statements concerning the employee's behaviour. Read attentively through each separate statement in turn. Think carefully and decide—on the basis of your observations of the employee concerned—whether the statement, on the whole, applies to him or her. If so, place a cross under "True". On the other hand, if you find that it does not apply, the cross must be placed under "Not true" instead. Crosses must be placed against all statements. However, indicate by writing *one* question mark if you feel that the assessment is uncertain in one or two cases and by writing *two* questions marks if the assessment approximates to pure guesswork.

The methods of making the observations and the filling in of the form were

also discussed at recurrent meetings between the author and the supervisors. Two to four weeks before the commencement of the criterion measurements, all the participants in the experiment were then observed by two or, in most cases, three supervisors. In evaluating the observations, I followed in each particular item the opinion expressed by two of the three observers. In the cases in which only two observers were available and their opinions differed, the one who had had the longest time to get to know the participant concerned was accepted as giving the decisive opinion.

I treated the observational data obtained in this way statistically with the aid of the Svarsum programme, which was placed at my disposal by the Institute of Education, University of Uppsala. On the basis of a content analysis of all items on the form, they were assigned to one of at most nine scales. After statements with a negative meaning had been turned, so that "True" was coded as "Not true" and vice versa, the frequency distribution, mean values, standard deviation and measurement of internal consistency (alpha) were calculated for each scale. Intercorrelations between the scales and point-biserial correlations between each item and the respective scale were also calculated.

By successive procedures of this kind, I gradually elaborated the maximum number of scales with satisfactory internal consistency, moderate intercorrelations and reasonable frequency distributions. It proved possible to design six such scales. I regard them in principle as measures of the characteristics in homogeneous and fairly well-defined spheres of behaviour, whose contents justify the following designations:

1. Emotional stability (24 items).
2. Ability to co-operate, with absence of tendencies to aggressiveness and domination (21 items).
3. Ability to communicate through speech (11 items).
4. Motivation to work (34 items).
5. Movement co-ordination and working pace (20 items).
6. Technical knowledge and skills (24 items).

These six scales, with all the items included, will be found in Appendix B:1 on pages 180–184. For each item, the number on the check list, the item-total correlations and the $p$ value in the total group have been given. The mean scores, standard deviations and reliability estimates (alphas) will be found in Table 6:1. Finally, the intercorrelations between the scores are given in Table 6:2. All the calculations refer to the somewhat reduced groups of men ($n$=40) and women ($n$=26) who took part in the testing programme presented later.

A close scrutiny of Table 6:1 will show that the reliability values are clearly satisfactory and thus that the scales to a large extent measure homogeneous variables. Furthermore the agreement between the observers was estimated by

*Table 6:1.* Number of items, means, standard deviations and estimates of reliability (alphas) in six sub-scales in the observation of working behaviour.

| Sub-scale no. | No. of items | Men, $n = 40$ | | | Women, $n = 26$ | | | $t$-values of sex differences |
|---|---|---|---|---|---|---|---|---|
| | | Means | Standard deviations | Alphas | Means | Standard deviations | Alphas | |
| 1 | 24 | 18.65 | 5.22 | .90 | 12.73 | 5.78 | .89 | 4.11 |
| 2 | 21 | 16.67 | 4.83 | .90 | 12.85 | 6.10 | .92 | 2.65 |
| 3 | 11 | 8.77 | 2.52 | .82 | 9.35 | 2.53 | .87 | −1.18 |
| 4 | 34 | 24.40 | 9.23 | .95 | 19.27 | 10.95 | .97 | 1.94 |
| 5 | 20 | 15.27 | 5.32 | .93 | 14.54 | 4.87 | .89 | 0.57 |
| 6 | 24 | 16.10 | 6.00 | .92 | 14.15 | 4.95 | .87 | 1.41 |

calculating the intra-class correlations. For this purpose I followed the procedure described by Guilford (1954), which Ebel (1951) had previously commented on and explained. The estimates of reliability thus obtained proved to be of the following magnitude:

1. Emotional stability .84
2. Ability to co-operate .84
3. Ability to communicate through speech .81
4. Motivation to work .84
5. Movement co-ordination and working pace .83
6. Technical knowledge and skills .75

While these values are on a somewhat lower level than the alpha coefficients reported in Table 6:1, they are highly satisfactory regarded as measurements of the reliability of systematic observations. They are examples of the fact that it is

*Table 6:2.* Intercorrelations between six sub-scales in the observation of working behaviour.

| Sub-scale No | | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | Men | .59 | .03 | .38 | .32 | .34 |
| | Women | .58 | .00 | .54 | .10 | .06 |
| 2 | Men | | −.22 | .24 | −.15 | −.08 |
| | Women | | −.12 | .60 | −.13 | .08 |
| 3 | Men | | | −.13 | .34 | .32 |
| | Women | | | −.16 | .08 | .34 |
| 4 | Men | | | | .47 | .59 |
| | Women | | | | .13 | .45 |
| 5 | Men | | | | | .81 |
| | Women | | | | | .35 |

possible to increase the reliability of such data by combining the judgements of several observers (Cronbach 1969).

The figures in Table 6:1 show further that the frequency distributions obtained are consistently negatively skewed. This is, of course, an expression of the fact that, in designing the observation form, I did not succeed—in spite of my conscious endeavours to do so—in wording the statements in such a way that a majority of them yielded $p$ values of about 0.50. Instead the majority of them yielded higher $p$ values. The tendency is so pronounced in the scale which measures ability to communicate through speech that I can only record that the differentiating capacity of this scale is manifestly unsatisfactory. In view of the fact that speech difficulties are comparatively common among the mentally retarded and often seem to constitute a clearly obstructive factor in their ability to adjust themselves to life (cf. Palmér 1968 (*b*)), however, I chose, in spite of the scale's imperfection, to include it as one of the other prediction variables, in the hope that it might conceivably yield tentative results that could be of interest. A further reason for this decision is the circumstance that the ability to communicate through speech shows negligible or low intercorrelations with other variables in the observation schedule and thus it may make a specific contribution to the total prediction ability in the observations.

In the other scales, the tendency to negatively skewed distributions is less pronounced and more tolerable. If we examine the intercorrelations between them and simultaneously bear in mind the logical analysis of their contents already performed, it turns out that the first and second scales seem to measure two fairly closely related variables of a clearly emotional stamp. The fifth and sixth scales also seem to measure two variables which are comparatively close to each other, but these refer to psychomotor and cognitive abilities. Finally, the fourth scale, which was found to measure motivation to work, occupies an intermediate position and shows moderate correlations with both the variables of an emotional stamp and the ability variables. However, there seems to be a certain sex difference in the latter respect, which means that "motivation to work" has more in common with the emotional variables in the female group, while, on the other hand, it is more closely related to the cognitive variables in the male group.

Finally, if we consider the mean values on the six scales, we find that the men display higher values than the women in all variables, except "ability to communicate through speech". The differences are most marked in the emotional variables, and here reach statistically significant levels, as shown by the calculated $t$-values. The differences within the cognitive variables are of lesser extent.

*Table 7.* Number of items, means, standard deviations and estimates of reliability (alphas) in two sub-scales of a five-degree rating schedule.

| Sub-scale no. | No. of items | Men, $n = 40$ | | | Women, $n = 26$ | | | $t$-values of sex differences |
|---|---|---|---|---|---|---|---|---|
| | | Means | Standard deviations | Alphas | Means | Standard deviations | Alphas | |
| 1 | 6 | 19.85 | 3.67 | .89 | 19.19 | 2.99 | .86 | .78 |
| 2 | 7 | 22.83 | 2.93 | .81 | 22.00 | 3.33 | .83 | 1.01 |

**C. Analysis of sub-scales in a five-degree rating schedule**

Alongside the general observation schedule of the check-list type just presented, I also designed a five-degree rating schedule. In doing this, I was able to build on the experience of others to a comparatively large extent (cf. e.g. Rubenowitz 1971) but I endeavoured nonetheless, by having repeated discussions with the supervisors and by holding experimental trials, to give the schedule a form that would make it well adapted to the special situation which exists at a sheltered workshop for the mentally retarded. The rating schedule contained 14 items. It was administered and treated on the same principles as the general observation schedule described above. I found it possible to distinguish two sub-scales, which I designated "rated work proficiency" and "rated adjustment to work" and which are reproduced in Appendix B:2. One item, which referred to a rating of the quantity of work, was correlated equally with both the sub-scales and was therefore omitted from them. In certain cases, I shall include it in the following studies and shall then regard it as a separate scale.

Table 7 shows the mean scores, standard deviations and reliability measures (alphas) in the two sub-scales. We can establish that the alpha coefficients even here display satisfactory values. The same applies to the intra-class correlations, which amount to 0.83 and 0.79. The intercorrelations between the two scales proved to be 0.49 among the men and 0.23 among the women. The frequency distributions obtained are approximately normal. Finally, in Table 7 we may note that the sex differences are small but nevertheless point in the same direction as in the general observation schedule of the check-list type.

# Chapter IV

# Criterion Variables: Selection and Results of Measurements

In the methodological discussion concerning prognostic investigations among the mentally retarded (cf. above, pp. 6 and 11 ff.), the criterion problems have been one of the main points. It would also seem to be beyond doubt that the defects in the criteria usually applied are to a considerable extent responsible for the contradictions and the inconsistencies in the results obtained. It is, of course, so much the more important to try and develop criterion variables which are less objectionable, at any rate from certain viewpoints. I considered this so important that in my main investigation I put in extensive and prolonged work on the measurement of the performances in certain tasks. However, before I did this, I obtained and used, in an earlier stage of the investigations, criteria which in principle are based on the subjective assessments of the workshop manager and the supervisors. In this connection, I shall first present these earlier criteria.

## 1. Ratings of Working Capacity

In all habilitation of the mentally retarded at different levels, a fundamental and realistic aim is to develop their potential working capacity as far as possible. As I pointed out in the introduction, this represents a narrower—and at the same time a more uniform—concept than social adjustment. Working capacity and adjustment to work are also the most important aim in the efforts at habilitation. If I possess a certain, though limited, working capacity, I can to a corresponding extent contribute to my own maintenance and feel a real personal satisfaction, whether I work on the open labour market or in more sheltered conditions.

As I also showed in the introduction, different individuals' degrees of working capacity may be the subject of observations and ratings, and to me it was a matter of trying to find suitable forms, in order to make use of these opportunities. In this connection, I defined working capacity as a quantitatively determined ability to carry out approved work performances in a sheltered workshop for the mentally retarded. On this basis, I designed three criterion variables, namely, the prognosis-group criterion, the percentage-group criterion and the wages-group criterion, which are presented in turn in the following subsections.

**A. The prognosis-group criterion**

This criterion variable was obtained by classifying the workers with respect to their suitability for placement in various work connections. In this case, I started with six placement alternatives, as follows:

(1) Institutional therapy.
(2) Work therapy.
(3) Industrial therapy.
(4) Sheltered workshop for the mentally retarded.
(5) Ordinary sheltered workshop.
(6) Open labour market.

The workshop manager gave a detailed account at the beginning of a meeting of the significance of these alternatives and also gave the supervisors who were present a written summary of this account. The supervisors were then assigned the task of marking separately on special forms their assessments of the respective worker's suitability on a scale which comprised six steps, corresponding to the above-mentioned placement alternatives. Thus, the figure 1 indicated that the assessor considered the worker suitable for placement in institutional therapy, and so on. As all the supervisors did not have an opportunity to observe all the workers, the number of assessors varied slightly from case to case. The lowest number was two and the highest eight, with an average of 6.6 assessors per person assessed. The mean value of the different assessments was calculated for each worker and constituted the final measurement.

I considered that there was good reason to assume that the variations in the scale designed in this way reflected variations in the working capacity. When the supervisors assessed a worker as suitable for one placement alternative or another, they had to start primarily with the ability shown in the different tasks in the sheltered workshop. In the instructions given to the assessors, however, it was pointed out that a certain amount of regard should also be paid to the degree of social function shown. The consequence of this is probably that the prognosis-group criterion is less to be regarded as a pure measurement of performance than the percentage-group criterion presented in the following subsection.

This may be one reason why the distributions of the prognosis-group criterion (cf. Fig. 6) do not show the same tendency to positive skewness which we shall find with regard to the percentage-group criterion. Instead they are approximately normal—or even exhibit a slight tendency to negative skewness—and differentiate satisfactorily. There is a certain sex difference, in that the men display a higher average value than the women, but it is not of such a magnitude as to be statistically significant.

I estimated the reliability of the criterion rating by calculating the intra-class correlation. As a result, the reliability of a single rater could be determined as .60. Whether we then applied the formula, originally proposed by Cureton

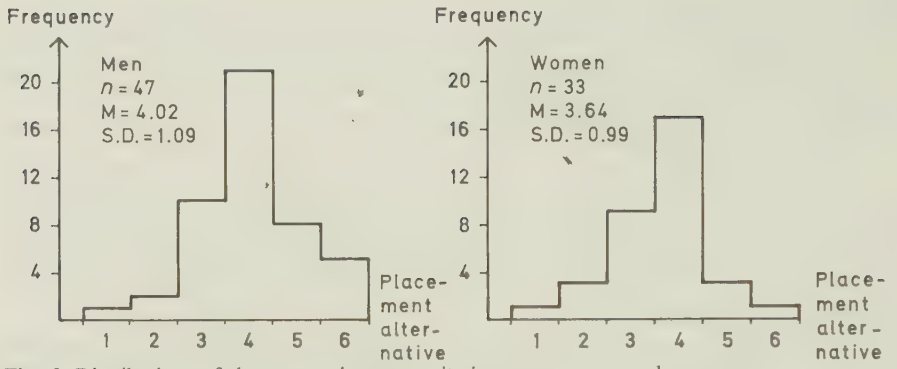


Fig. 6. Distributions of the prognosis-group criterion among men and women.

$$r_{kk} = \frac{V_p - V_e}{V_p}$$

or the Spearman-Brown formula for predicting a measure $k$ times as long, we found that the reliability of the average rating which the prognosis-group variable represents amounted to .91. Thus, the observation made in other connections—that satisfactory reliability can be achieved by combining the ratings of several assessors—was confirmed in this case.

**B. The percentage-group criterion**

This criterion variable got its name from the fact that we asked the supervisors to make an assessment of the percentage work contribution which the respective worker was able to make in relation to the normal worker. The supervisors were set this task in connection with one of the regular conferences at the workshop. In the instructions, the workshop manager requested them to recall carefully the observations they had previously made of the workers' performances in different tasks and to base their assessment on these previous observations as far as possible. Independently of each other, the supervisors had then to fill in a form, which contained a seven-degree scale, as follows:

| Scale step | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Percentage | 1-15 | 16-30 | 31-45 | 45-60 | 61-75 | 76-90 | <90 |

This criterion rating was carried out in two rounds. On the first occasion, 64 workers were assessed by eight supervisors, who then rated—in accordance with thier varying knowledge—61, 55, 45, 45, 42, 40, 26 and 21 individuals respectively. On the second occasion, the remaining 16 workers were assessed. Nine supervisors took part in this and rated 16, 16, 16, 13, 13, 10, 9, 8 and 6 individuals

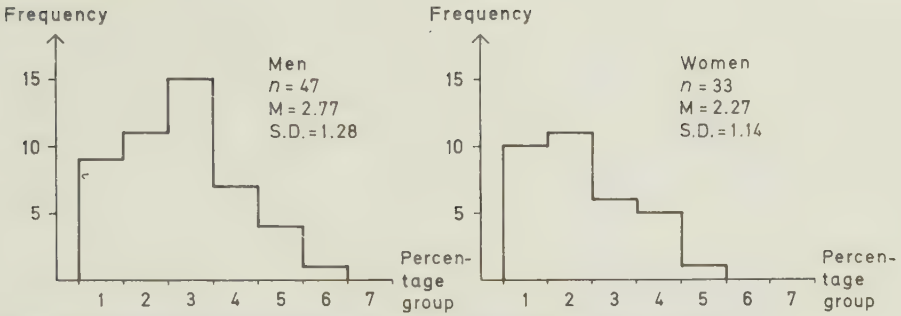


Fig. 7. Distributions of the percentage-group criterion among men and women.

respectively. When the average assessment figures had been worked out, the criterion measurement was found to display the distribution between men and women shown in Fig. 7.

This criterion variable also differentiates satisfactorily. A certain tendency to positive skewness appears, not unexpectedly, in both men and women. We also find—just as in the prognosis-group criterion—that the men achieve higher values than the women, though the difference is not statistically significant.

As regards the percentage-group criterion, I may point out, in conclusion, that the conditions for the assessment were good. All the supervisors who took part had been able to follow the workers' performances at the sheltered workshop daily for at least 3 months, and in most cases the observation period was in fact much longer. A calculation of the intra-class correlation also confirmed that the reliability is satisfactory. Applying the same procedure as in the case of the prognosis-group criterion, I found that the reliability amounted to .87. (I was able to determine the reliability of a single rater as .54; in this case also, the formula proposed by Cureton and the Spearman-Brown formula yielded identically the same result.)

### C. The wages-group criterion

This variable was calculated on the basis of the workers' placings in the wages grades. These placings were checked and fixed at least once a month at meetings of all the supervisors under the chairmanship of the workshop manager. It was clearly stated that decisions about placings in the wages grades were to be based solely on the workers' performances and therefore there is also reason to expect that the variation of the wages grades will reflect fairly well the variation of working capacity in those employed at the workshop.

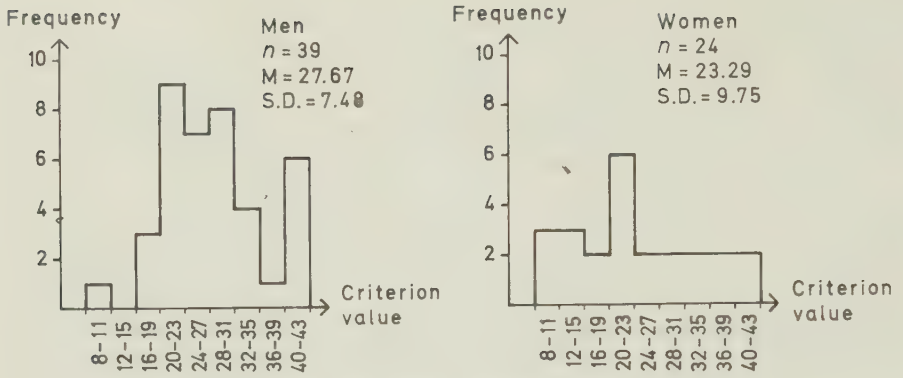



Fig. 8. Distributions of the wages-group criterion among men and women.

The wages plan included four wages grades, within which the wages varied between 1 Swedish crown per hour and 1.45 crowns per hour. Thus, the amount was almost symbolic, but all the same the workers seemed to attach considerable importance to their placings and the workshop management carefully examined all decisions on the question. The criterion measurement was designed by calculating the average of the wages grades fixed for the respective worker during 10 consecutive months. There was an obvious advantage in being able to base the measurement on observations over such a long period. Temporary fluctuations in the work performances were levelled out and the workers had opportunities to demonstrate their working capacities in different tasks. In the few cases in which the workers had not been at the workshop for so long a period as 10 months, the measurement had to be based on a shorter period (at least 3 months). The average figure calculated was multiplied by 10, in order to avoid the use of a decimal point. In connection with the preparation of the first report on the research project, frequency distributions of the criterion measurement calculated in this way were drawn up. They comprised 39 men and 24 women and will be found in Fig. 8.

The histograms show that the wages criterion seems to differentiate in a satisfactory fashion between both male and female workers. A fairly obvious sex difference appears, moving in the same direction as in the other two criterion variables. However, it does not achieve full statistical significance here either.

The design procedure just described would seem to have resulted in the criterion fulfilling reasonable requirements for reliability. However, it proved to be impossible to obtain values with the same content for the workers who were employed at the workshop during the later part of the experimental period, since a noticeable change in the grounds of assessment gradually appeared in connection with the wages-grade placings. For various reasons, the workshop manage-

*Table 8.* Correlations between three criterion variables.

| | | Percentage group | | Wages group | |
|---|---|---|---|---|---|
| | | Men $n = 47$ | Women $n = 33$ | Men $n = 39$ | Women $n = 24$ |
| Prognosis group | Men $n = 47$ | .74 | | .76 | |
| | Women $n = 33$ | | .61 | | .68 |
| Percentage group | Men $n = 47$ | | | .70 | |
| | Women $n = 33$ | | | | .80 |

ment became more generous in its placings in high wages grades, and this led to a shift in the proportional distribution in different grades. On this account, I was compelled to refrain from calculating criterion values of this kind for eight men and nine women. Because of this, I shall use this criterion only to a limited extent in the further treatment of the material.

**D. Correlations between the criterion variables**

The three criterion variables presented above are accordingly all careful ratings of the working capacity shown in a sheltered workshop. As is to be expected, they also display substantial or high intercorrelations, as is shown in greater detail in Table 8. I shall try to analyse these intercorrelations in more detail after I have described the other criterion variables.

## 2. The Measurement of Work Performances

**A. The selection of tasks**

In the introduction, I referred to investigations by Schulman and Elkin (cf. pp. 14 ff.) as examples of attempts in previous prediction studies to obtain completely objective criteria by measuring the work performances of the mentally retarded. The reported results are valuable and interesting, but their scope is obviously comparatively limited. It is time-consuming to make such measurements under realistic conditions. Schulman allowed his experimental subjects to work only 3 hours per day for a 5-day week, during which they were engaged in several different tasks, while Elkin allowed his experimental subjects to work on each of four tasks for 45 minutes for 6 consecutive days. The total time which the experimental subjects devoted to each task was consequently fairly short in both cases, and therefore it is to be feared that the results of the measurements refer to

performances at a comparatively early stage of a training course whose appearance is unknown.

One of the principal aims in the planning of my own criterion measurements was to extend the working time and to try and carry out the measurements also in other respects under conditions characteristic of a normal working environment. From this point of view, it was extraordinarily important that I was able to locate the investigations in a workshop which had come into existence for the particular purpose of facilitating the scientific investigation of problems connected with the habilitation of mentally retarded persons (cf. above, p. 20).

Another basic aim was that I should select or design tasks which permitted to the utmost possible extent the demonstration and analysis of working capacity in mentally retarded persons. I was able to obtain several ideas on this point from observations and general experience in the Uppsala workshop. However, the inventory of sheltered workshops for the mentally retarded in Sweden which was made in 1967 also yielded a great deal of guidance. In my account of this inventory (Palmér 1968*a*), there is a fairly complete list of the tasks which were carried out at that time in these workshops, together with the workshop managers' assessments of their suitability for the clientele.

In these different ways, I obtained a general idea of the nature of the tasks which should be considered. On the basis of this, I then tried to produce variations as regarded such factors as the initial material (metal, wood, cloth, plastics), the length of the work cycle, the degree of employment, and the requirements as regarded precision and co-ordination of movements etc. These considerations, in which, of course, questions of cost and other practical assessments necessarily also played an essential part, resulted in the selection of eight tasks. Three of these tasks, viz. the pressing of eyelets, the cutting of threads in nuts and the turning of holders, were regular piece-work tasks in the engineering section. Two other tasks were carried out in an assembly section. One of these was comparatively simple and could be performed rapidly; it involved assembling a bracket for an exhaust pipe and also formed a part of the regular workshop production. The other was purposely designed for the present investigation and consisted of the somewhat more complicated assembly and wiring of an electrical panel. In addition, there were a further three specially designed or experimental tasks, viz. the making of a wooden frame, the embroidering of cat motifs and the packing of blocks (for detailed descriptions, see Appendix C on p. 186 ff.).

For each of these tasks, special workplaces were designed, in which the principles of economy of motion and biotechnology were applied. Complete Methods-Time Measurement (MTM) analyses were carried out in close connection with this work (cf. Hasselqivst *et al.* 1967, Gold 1973). This gave me more reliable opportunities of assessing the nature of the demands which the different

tasks made on the workers. Furthermore, the MTM analysis facilitated the calculation of standard values, on the basis of which certain comparisons may be made between the performances of the normal worker and those of the mentally retarded.

On the basis of the MTM analysis and in close consultation with the workshop manager and the supervisors, I drew up written instructions for the various tasks (Palmér *et al.* 1970). In doing so, I endeavoured to apply, with as much insight as possible, the current principles of teaching methodology. In my instructions to the supervisors, I pointed out in particular that, as instructors, they should endeavour to instil calmness and confidence into the workers and should themselves adopt a positive and encouraging attitude, particularly when giving instructions and at the beginning of a working period. By holding recurrent discussions, I tried to co-ordinate the supervisors' conduct and to make it uniform also in other respects. In this way, I wished to reduce the risk of the supervisor factor becoming a source of error, in the same sense as the teacher factor has been a source of error in several educational investigations.

**B. The combination of the tasks to form the testing programme**

When the eight tasks had been selected and the necessary workplaces had been built, we set out the workplaces in the different sections of the workshop to which they naturally belonged. This step was a stage in the endeavour to co-ordinate, as smoothly as possible, an extensive testing programme with the daily workshop routine. In this connection, I also had to make decisions on a series of fundamental practical problems of a heterogeneous nature. The most important of these concerned the length of the testing periods and the sequence of the tasks.

As far as the length of the testing periods was concerned, I realized that a few hours' or even two days' work on the respective tasks would not be sufficient. In that case, I would obtain too unreliable a measurement of performance with regard to the fluctuations in the work cycle which were to be expected. Even more serious was the risk that in these circumstances I would only be able to show the very beginning of a training course and that I would, on the whole, not have satisfactory opportunities to observe the working behaviour in real-life situations. On the other hand, I had to note that in its entirety the investigation would take too long a time and be too expensive, if I extended the working periods to cover more than a week per task. When, finally, I considered the practical resources available for simultaneously giving instructions, making observations and checking results with reference to eight workers, I found it necessary to abandon Friday in the working week and consequently to limit the length of the testing periods to 4 days.

Another question was concerned with the sequence of the eight tasks. Should I present them to all the workers in one and the same order or should I use a

system of rotation, which would mean that each task was presented just as often as nos. 1, 2, 3, etc. The answer to this question had to do with the aim of the investigation. If I intended primarily to test the individuals' working capacity as a whole and to secure a means of predicting this on the basis of the test results obtained, I should rather apply the first alternative. I should thereby best fulfil the basic requirement in all test methodology—that the different individuals should, as far as possible, be placed in comparable situations. If, on the other hand, I regarded it as most important to analyse the structure of the working capacity of the mentally retarded and to compare it with that of workers of normal intelligence, the application of the first alternative would be questionable. If the eight tasks are presented in the same order throughout, transfer effects may arise and may improve the performances in the tasks which come towards the end of the series. This makes impossible true comparisons between the performances in tasks nos. 1 and 8, for example. Thus, here I would have a reason for applying instead the second alternative with a system of rotation.

However, the second alternative had a disadvantage associated with it, in that I wished to make comparisons between individual performances in the different tasks. For example, an individual may have been entrusted with the packing task as the first task in his series, while another individual has encountered it as his last task. If transfer effects occur, comparisons between the performances of the two individuals will not be entirely correct. The same applies to intra-individual comparisons.

Since I rather wished to achieve several aims in the investigation, I was in a dilemma here. I solved the problem by deciding on the second alternative with a system of rotation. The not unimportant circumstance that otherwise I would have lost 7 weeks at the start argued in favour of this course. Accordingly, in view of the part which transfer effects might conceivably play, I also considered it to be desirable to try, on the one hand, to obtain a measurement of the strength of these possible effects and, on the other, to plan the investigation in such a way as to allow them as little scope as possible in all circumstances.

In order to achieve the first of the above-mentioned aims, I divided the eight tasks into two comparable groups of four each. Cutting threads, turning, electrical-panel assembly and packing were assigned to the first group, and pressing eyelets, wooden-frame-making, sewing and bracket assembly to the second. Furthermore, I likewise divided the experimental subjects into two groups, which were made as alike as possible by being matched as regarded sex, length of preliminary training, previously estimated working capacity and performance in tests of psychomotor ability. During the first working period, one of these groups of experimental subjects had to apply itself to and complete the work on the thread-cutting machine, lathe, electrical-panel assembly and packing. Only then did I allow them to pass on for the second working period to the excenter press,

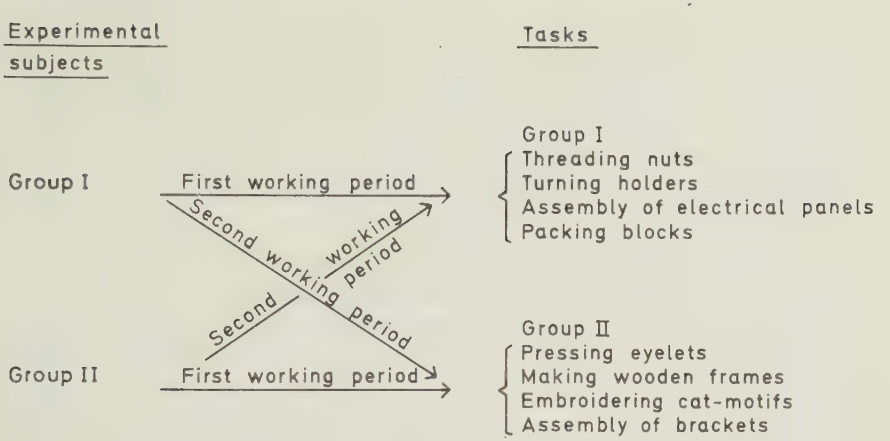



Fig. 9. Sketch of the arrangement for studying transfer effects.

wooden-frame-making, sewing and bracket assembly. The second group of experimental subjects had to go in the reverse order and thus complete the tasks on the excenter press, wooden-frame-making, sewing and bracket assembly, before they were allowed to pass on to the thread-cutting machine, lathe, electrical-panel assembly and packing. The arrangement is shown in the sketch in Fig. 9.

If transfer effects occur now, this should appear as a difference in the performances between the two groups of experimental subjects. Thus, group I should display a somewhat worse performance in the first group of tasks and a somewhat better performance in the second group of tasks, in comparison with the second group of experimental subjects.

In order to achieve the second aim—to restrict, if possible, the scope for any possible transfer effects—I had recourse to an arrangement in which the testing periods were dispersed in time. Between two such periods, the worker was allowed to return to the ordinary workshop work for a week. This meant that the individual experimental subject was engaged in the investigation every other week for a total experimental period of 16 weeks. This arrangement had both advantages and disadvantages from the practical point of view. It required that I should have each experimental subject at my disposal for a fairly long period of time in all, which sometimes presented difficulties. However, the advantages would seem to have predominated. Amongst other things, it was possible to transfer visits to the doctor and the dentist and other unavoidable actions to the intermediate weeks.

The results of the control study on the extent of possible transfer effects are reported in Appendix D on pages 216—218. There it is suggested that positive transfer effects only influenced the results of the criterion measurements to a very

limited degree. Signs of such influence only appeared in "making wooden frames". However, one cannot entirely ignore the possibility that certain negative effects, caused by incipient fatigue or waning motivation may, on the other hand, have resulted in a small decrease of the performances in some of the other tasks. Yet it is in any case a question of minor effects, which we may well assume did not contribute to any crucial extent to reduce the homogeneity of criteria or to make individual comparisons impossible.

**C. The gradual implementation of the testing programme**

With the guidance of the above decisions regarding the length of the testing periods and the order of the tasks, I drew up a detailed plan for the gradual implementation of the testing programme. The available resources, as regarded money and premises, did not permit of the simultaneous testing of more than eight workers. However, in view of the fact that—in accordance with the dispersal in time of the testing periods, as explained above—these workers would only take part in the investigation every other week, I had to select, from the beginning, twice that number or 16 workers, whom I regarded as my first working group. When they had completed their participation in the testing programme after altogether 16 weeks, my task was to select a second working group of 16 workers in the same way, and so on. In all, I was able to allow four such working groups to undergo the programme from November 1968 to June 1970.

With regard, amongst other things, to the circumstance that in the initial situation I could not know for certain how long I would have the financial resources to complete the investigation programme without restrictions, I found it advisable to select the workers in such a way that each working group was, as far as possible, representative of the whole experimental group. In any event, a matching procedure was necessary, in order that I might be in a position to measure any possible transfer effects, in the way mentioned above.

I have already mentioned that I intended to use as matching variables (1) sex, (2) length of preliminary training, (3) previously rated working capacity, and (4) performances in tests of psychomotor ability. As concerned the sex variable, I let each working group consist of 10 men and six women, which corresponds fairly well to the proportional distribution in the whole experimental group which I reported above. I was only incompletely able to check the length of the preliminary training. I had to be content with regarding it as being synonymous with the period of employment at the Uppsala workshop and with constituting the four working groups so that from this point of view they were as like each other as possible. Table 9 shows that I had only moderate success in this endeavour. Thus, I was unable to prevent the length of the employment period from becoming considerably shorter in the fourth working group and the range of this variable from increasing appreciably from the first to the last group.

*Table 9.* The four working groups and their values in the matching variables.

| Working group | Date of testing | Period of employment in months | | Prognosis group | | Percentage group | | Psychomotor ability | |
|---|---|---|---|---|---|---|---|---|---|
| | | M | S.D. | M | S.D. | M | S.D. | M | S.D. |
| I. $n = 16$ | 25.11.68–11.4.69 | 20 | 1.82 | 3.94 | 1.14 | 2.56 | 1.54 | 15.88 | 6.85 |
| II. $n = 16$ | 14.4.69–11.9.69 | 19 | 2.71 | 3.75 | 0.66 | 2.63 | 1.05 | 14.69 | 5.96 |
| III. $n = 16$ | 15.9.69–22.1.70 | 17 | 5.82 | 4.00 | 0.87 | 2.63 | 1.11 | 15.88 | 4.59 |
| IV. $n = 16$ | 26.1.70–4.6.70 | 11 | 8.84 | 3.94 | 0.97 | 2.50 | 1.12 | 15.31 | 6.18 |
| Total $n = 64$ | 25.11.68–4.6.70 | 17 | 6.56 | 3.91 | 0.93 | 2.58 | 1.22 | 15.44 | 5.97 |

However, as regarded the remaining—and probably more important—matching variables, I had better success. Before they underwent the testing programme, the working abilities of all the participants had been rated in accordance with the procedure which was described above and which resulted in the prognosis-group and percentage-group criteria.

The values inserted in Table 9 show that the four working groups are to a large extent comparable in these respects, though certain differences appear as regards the dispersion of the values. Finally, the performances in tests of psychomotor ability were expressed as the totals of the stanine scores in four tests, viz. Crawford Screws, Crawford Pins, the O'Connor Test and Wire-design Copying (cf. Kebbon 1965). As Table 9 shows, the similarity between the four working groups in this variable is very obvious.

After four working groups or altogether 64 workers had undergone the testing programme, the investigation had to be broken off, in consequence of far-reaching organizational changes in the running of the workshop. This meant that seven men and nine women out of the total experimental group, whom I described in a previous section, never had an opportunity to demonstrate their ability in experimental tasks. However, owing to the matching procedure described above, I was able to anticipate that the group who had undergone the testing programme would be comparable in all essential respects with the total experimental group. The figures in Table 10 confirm that this is the case. If we analyse the differences between the group who underwent the testing programme and the drop-out group, we shall find that these differences do not achieve a statistically significant magnitude, as regards either age, IQ or scores in the two rating criteria.

*Table 10.* Comparison between the characteristics of the whole experimental group (T) and those of the group which underwent the testing programme (P).

| | Men | | Women | | Educable | | Non-educable | | Age | | IQ | | Prognosis group | | Percentage group | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % | M | S.D. | M | S.D. | M | S.D. | M | S.D. |
| Group T $n = 80$ | 47 | 58.7 | 33 | 41.3 | 52 | 65 | 28 | 35 | 25.3 | 4.53 | 53.3 | 10.22 | 3.86 | 1.06 | 2.56 | 1.25 |
| Group P $n = 64$ | 40 | 62.5 | 24 | 37.5 | 40 | 62.5 | 24 | 37.5 | 25.4 | 4.84 | 52.8 | 9.73 | 3.91 | 0.93 | 2.58 | 1.22 |
| Group T – P $n = 16$ | 7 | 43.8 | 9 | 56.2 | 12 | 75 | 4 | 25 | 24.6 | 2.93 | 55.1 | 11.81 | 3.69 | 1.45 | 2.50 | 1.32 |

*Table 11.* Average percentages of approved products in eight tasks.

| | Pressing eyelets | Cutting threads in nuts | Turning holders | Assembling brackets | Assembling electrical panels | Making wooden frames | Embroidering cat motifs | Packing blocks |
|---|---|---|---|---|---|---|---|---|
| Men $n = 40$ | 95 | 99 | 74 | 99 | 87 | 73 | 39 | 78 |
| Women $n = 24$ | 93 | 99 | 60 | 99 | 77 | 54 | 55 | 76 |

The division into the categories of educable and non-educable was not altered to any decisive extent either. On the other hand, one cannot entirely disregard the fact that the proportion of women, which was comparatively low even in the total group, was subject to a further reduction in this connection. Consequently, the prerequisite conditions for making comparisons between men and women must be said to be more limited than before. In other respects, however, the possibilities of generalizing would not seem to have undergone any marked deterioration in this part of the investigation, in spite of the above-mentioned reduction in the number of the experimental subjects.

**D. The nature of the tasks and the measurements of qualitative performance**

*Requirements for the approval of finished products*

Before the testing programme could begin, I had to lay down rules for evaluating the performances. In doing this, the most important thing was to determine, with sufficient accuracy, the requirements which must be fulfilled for a performance to be approved. I soon became convinced that in this connection I should in principle adhere to the corresponding conditions on the open labour market. In other words, only such performances should be approved as are accepted there as satisfactory.

It was so much the more natural in this context to adhere to the conditions on the open market as four of the tasks were included in the workshop's regular production, which was regularly checked and inspected before delivery to the client. However, we tried to apply the same fundamental view also in other tasks. What this meant in concrete terms and in detail is described in Appendix C, which also mentions how, in connection with the checking and inspection, we classified different types of fault and noted the frequencies with which they occurred.

*The degree of difficulty of the tasks*

As an expression of the degree of difficulty of the tasks, I first calculated the average percentage of approved products for men and women during the last 2 days of the total working period.

These average percentages proved to vary as shown in Table 11. We find here that two of the tasks, viz. the cutting of threads in nuts and the assembling of brackets, differ from the rest in that the percentages of rejections are quite negligible. Thus, there is reason to regard them as the easiest tasks from the qualitative point of view. However, they are closely followed by the pressing of eyelets, which did not present the participants in the experiment with any great difficulties either. The assembling of electrical panels, the packing of blocks and the turning of holders may be considered to form an intermediate group from this point of view. Finally, it is evident that the making of wooden frames gave the

women in particular marked difficulties and that embroidering cat motifs was the most difficult of all the tasks. To summarize, the figures in Table 11 indicate that I had some success in my efforts to select tasks which represented different levels of difficulty.

This ranking order of the eight tasks with respect to their degrees of difficulty seems fairly reasonable, if we take into consideration the detailed descriptions in Appendix C. In these descriptions, I have endeavoured to state—in addition to accounts of the work cycles and the associated patterns of movement—in concrete terms the critical operations which are included to a greater or lesser extent in the respective tasks. I have also reported a technical assessment of the requirements which these tasks make as regards persistence, psychomotor skill, speed of movement, concentration, strength, ability to co-ordinate and ability to use tools and to follow patterns of movement.

Without making a detailed analysis of these data, we may note that the work cycles in the cutting of threads in nuts, the assembling of brackets and the pressing of eyelets are short and consist of simple patterns of movement and that the perceptual discriminations which the workers have to make in these tasks are comparatively few and easily managed. The fact that the assembling of brackets and the pressing of eyelets are considered to make great demands as regards persistence, in consequence of their monotonous character, did not prevent qualitatively good outputs from being produced, a circumstance which is well worth noting. In the tasks which were assigned to an intermediate group from the point of view of difficulty, viz. the assembling of electrical panels, the packing of blocks and the turning of holders, the patterns of movement are somewhat more complicated and moreover elements of perceptual discrimination are included in greater numbers. According to the technical assessment, they also make greater demands on concentration. In the two most difficult tasks, viz. the making of wooden frames and the embroidering of cat motifs, there are also several elements of perceptual discrimination, but here the worker also has to learn longer sequences of movements and has to be able to master certain subsidiary movements which are characterized by precision and variation. Accordingly, the technical assessment also states that these tasks make great demands on concentration.

What may be surprising, when we compare in this way the calculations of the percentage of approved products with the detailed descriptions of subsidiary operations included in the tasks and the assessments of requirements made in connection with them, is that the participants in the experiment were able to put up such comparatively good performances in the assembling of electrical panels. This task also requires, to no small extent, that the worker shall learn a long sequence of movements and shall be able to master precise and varied subsidiary movements, which in their turn make considerable demands on concentration,

speed of movement and psychomotor skill. I have not been able to develop any definite hypothesis which might explain this circumstance and regard it primarily as a manifestation of the fact that our knowledge of the relation between the nature of a task and its degree of difficulty is inadequate.

*Qualitative development of performance*

The nature of the task will also be decisive as regards the extent to which training results in qualitatively improved outputs. I have tried to illustrate this fact by calculating the percentage of approved products made during eight consecutive working periods. Here it should be recalled that the participants in the experiment were engaged in each task for altogether 4 days and that, in doing so, they kept the workshop's normal working hours, which extended from 08.00 in the morning to 16.40 in the afternoon, with the usual lunch and coffee breaks. These eight consecutive periods consisted of the mornings and afternoons of the whole working period. The results of the calculations will be found in Table 12.

In examining this table from the above-mentioned point of view, we should, on the whole, disregard the first two working periods, since the instructors were then actively teaching the workers and giving them guidance. If, therefore, we concentrate our attention on the percentage figures in the next six columns, we find that

*Table 12.* The qualitative development of work performance, expressed as a percentage of the approved products during each of eight consecutive working periods.

| Working period no. | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| Pressing eyelets | Men | $n = 40$ | 96 | 96 | 96 | 95 | 96 | 95 | 95 | 95 |
| | Women | $n = 24$ | 93 | 94 | 94 | 94 | 93 | 93 | 94 | 94 |
| Cutting threads in nuts | Men | $n = 40$ | 98 | 98 | 99 | 98 | 99 | 97 | 99 | 99 |
| | Women | $n = 24$ | 98 | 97 | 99 | 99 | 100 | 99 | 99 | 99 |
| Turning holders | Men | $n = 30$ | 69 | 60 | 65 | 62 | 70 | 71 | 76 | 75 |
| | Women | $n = 18$ | 52 | 46 | 45 | 42 | 53 | 60 | 58 | 64 |
| Assembling brackets | Men | $n = 40$ | 100 | 100 | 100 | 99 | 99 | 99 | 99 | 99 |
| | Women | $n = 24$ | 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| Assembling electrical panels | Men | $n = 40$ | 95 | 86 | 91 | 89 | 90 | 89 | 87 | 86 |
| | Women | $n = 24$ | 98 | 85 | 78 | 74 | 78 | 72 | 84 | 85 |
| Making wooden frames | Men | $n = 40$ | 77 | 72 | 68 | 68 | 69 | 74 | 74 | 74 |
| | Women | $n = 24$ | 55 | 50 | 46 | 46 | 53 | 49 | 64 | 52 |
| Embroidering cat motifs | Men | $n = 40$ | 49 | 38 | 46 | 37 | 45 | 39 | 42 | 34 |
| | Women | $n = 24$ | 66 | 56 | 65 | 60 | 62 | 60 | 54 | 51 |
| Packing blocks | Men | $n = 40$ | 73 | 72 | 76 | 75 | 76 | 76 | 77 | 77 |
| | Women | $n = 24$ | 67 | 68 | 77 | 78 | 77 | 74 | 77 | 76 |

the cutting of threads in nuts, the assembling of brackets and the pressing of eyelets show a high and even development of performance. Of course, no training effects of a qualitative nature can appear in tasks which are so easy that the participants in the experiment have already mastered them at the start.

No such training effect can be distinguished in the packing of blocks or the embroidering of cat motifs either. In the former task, the qualitative development of the performance is obviously even. Almost 25 % of the production is rejected during the whole working period, and the difficulties which cause this evidently cannot be overcome, at any rate not in so short a training period as 4 days. In the sewing task, the development of the performance is more uneven. In both men and women, there appears, in actual fact, a tendency to performances that are qualitatively worse as the working week progresses. Moreover, we may note that the quality of the work appears to be lower during the afternoons than during the mornings. It may very well be imagined that both these circumstances are due to the fact that the task makes such great demands on the workers—chiefly as regards concentration and psychomotor ability—and that they therefore become exhausted even after short periods of work.

We find clear signs of qualitative improvements of the outputs with increased training only in the turning of holders and the making of wooden frames and—as far as the women are concerned—the assembling of electrical panels. Thus, it seems as if even 4 days' training results in greater ability in the workers to meet the demands which these tasks make on them, as regards the learning of sequences of movements, the use of tools, etc. There may be several reasons why the men do not show signs of improvements of this nature in the assembling of electrical panels. However, the most likely explanation is that they had received more training in similar jobs than the women before the investigation and that they had thereby already acquired certain basic skills which the women acquired only when the investigation began.

*Measurement of qualitative performance*

In this connection, I finally also examined the distributions of the individual values with regard to the percentage of approved products. As may be expected, the appearance of these distributions varies markedly with the degree of difficulty of the tasks. In the easier tasks, the great majority of individuals have the maximum value of 100 and the distributions are consequently extremely skewed. As the degree of difficulty increases, the ability of the tasks to differentiate in the given respect also increases and the distributions become more extensive. In this situation, I found it meaningful to calculate the average percentage of approved products made by the participants in the experiment, as a composite measurement of their qualitative performance in the eight tasks. Fig. 10 shows how these values are distributed.

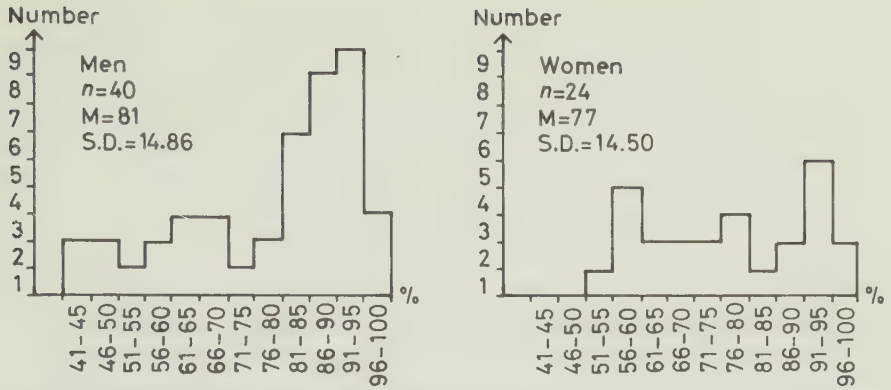



Fig. 10. Distributions of the average percentage of approved products in eight tasks among men and women.

We find here that the variable calculated in this way differentiates fairly satisfactorily in both men and women. In the male group, a tendency to negative skewness appears in the distribution, while in the female group a tendency to bimodality is discerned. The men's mean value is higher than the women's, but the difference is not so great as to be statistically significant.

**E. Quantitative development of performance and measurement of quantitative performance in the eight tasks**

The measurement of qualitative performance presented in the preceding section is an important criterion of working capacity. However, it is of even greater importance to calculate the measurement of quantitative performance, as it appears in the number of approved products per unit of time in the respective tasks. Indeed it is chiefly the quantity of approved products which is primarily decisive of an individual's chances of asserting himself on the labour market.

*Quantity and quality*

Here it may be asked whether we ought not also, in one way or another, to include the number of rejected products or the quantity of "waste" in a criterion of working capacity. This would be justified if it was feared that the qualitative capacity varied fairly independently of the quantitative. In order to shed light on this question, I have examined how the numbers of approved and rejected products are related to each other in the different tasks. In this connection, I was able to note that there is a fairly marked correlation between them, which means that the number of rejected products tends to decline at the same rate as the number of approved products increases. As a logical consequence of this, I also found that the product-moment coefficient of the correlation between the com-

posite measurement of quantitative performance (which I shall present later on) and the measurement of qualitative performance (which I have just described) reaches in the whole group a magnitude of .69 (cf. Table 21 on page 93). Fig. 11 is a correlation diagram which illustrates in greater detail the nature of this correlation. What immediately strikes us here is that none of the individuals in the experimental group falls above the diagonal line. This means that workers whose quantitative outputs are relatively high also satisfy the qualitative requirements well. This may also be interpreted as meaning that tendencies to lack of accuracy or general negligence at work seldom occur. However, the diagram also shows that some workers whose outputs are low from the quantitative point of view also display a high percentage of approved products. This circumstance, which is chiefly responsible for the fact that the calculated correlation coefficient is not greater than .69, results in the ranges of the quality measurement being

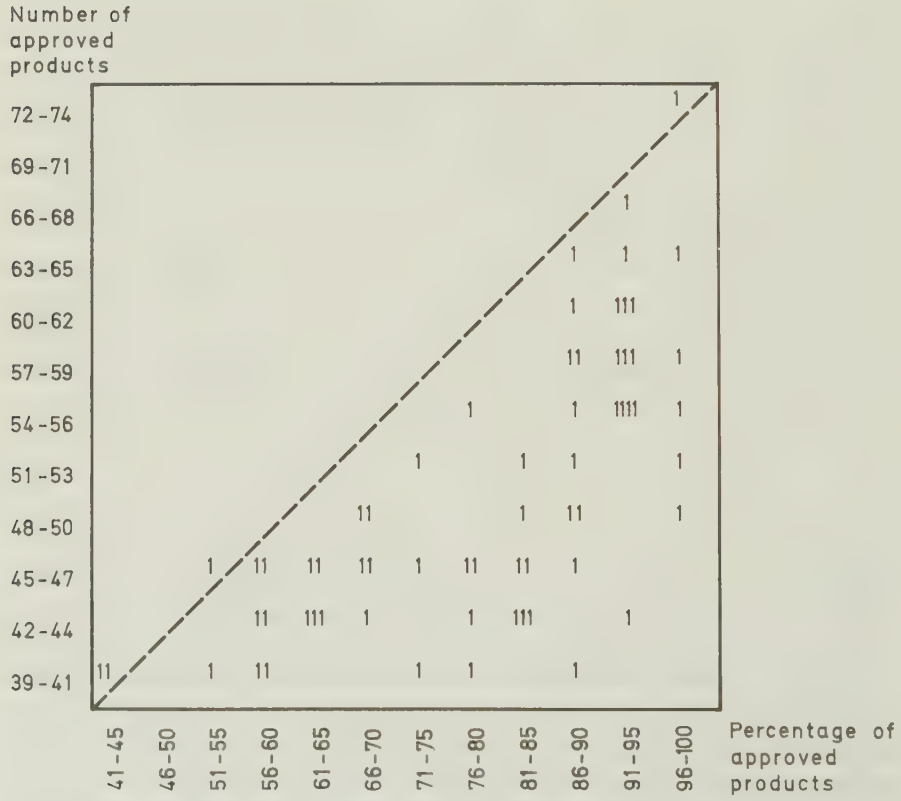



Fig. 11. Correlation diagram, showing the correlation between quantitative performances (number of approved products) and qualitative performances (percentage of approved products).

comparatively large at low levels of quantity. It is, of course, important to take due account of this, whenever one wishes, for example, in giving guidance, to describe in greater detail the working behaviour of the individual employee. However, in this connection I did not find sufficient reason for trying to design a criterion variable in which the dimensions of quantity and quality could be weighted together.

*Quantitative development of performance*

In order to accomplish my aim to measure the quantitative performances, I first analysed the quantitative development of performance in each task, by calculating the number of approved products per hour during eight consecutive working periods. As I have already mentioned, these periods consisted of the mornings and afternoons of the four working days.

One detail connected with the calculations of these successive measurements of output is worth mentioning here. A morning period comprised 3 hours and 40 minutes and an afternoon period 3 hours and 20 minutes. When I divided the number of approved products by these measurements of time, I made deductions for involuntary interruptions, for example, visits to the doctor and similar matters, for which the worker had received special permission. If, on the other hand, he stopped work of his own accord for a varying period, I made no corresponding deduction,, as I regarded this behaviour as a manifestation of a correspondingly reduced working capacity. The same principle was followed in the calculations of the total measurement, which are presented later on.

The quantitative development of performance determined in this way is reported numerically in Table 13. I have also illustrated it in diagrams, which will be found in connection with the detailed descriptions of the tasks in Appendix C. What I am primarily interested in here is really the direction and form of the development. Is there an increase or a decrease in the outputs from one working period to another and are there any systematic tendencies in these changes?

A closer study of the table and the diagrams shows that, as a rule, the outputs have a tendency to increase from the first working day to the last. Of course, I consider that there is good reason to interpret this as the effects of training. There is only a single departure from this rule and it refers to the outputs of women in embroidering cat motifs, which show a slight reduction instead of an increase. Several circumstances may have co-operated to ensure the non-appearance of training effects in this case. Thus, the task invited very active guidance on the part of the instructor during the first working period, which resulted in a comparatively good initial performance. Furthermore, there was no doubt that the female experimental subjects had, as a rule, received, during their childhood and adolescence, in accordance with the traditional cultural pattern, considerable training in sewing work, the consequence of which was that they already

*Table 13.* The quantitative development of work performances, expressed as the average number of approved products per hour during each of eight consecutive working periods.

| Working period no. | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Pressing eyelets | Men $n = 40$ | 246 | 265 | 280 | 273 | 286 | 285 | 314 | 307 |
| | Women $n = 24$ | 207 | 248 | 259 | 245 | 282 | 268 | 296 | 264 |
| Cutting threads in nuts | Men $n = 40$ | 188 | 181 | 199 | 183 | 203 | 194 | 205 | 192 |
| | Women $n = 24$ | 150 | 152 | 169 | 154 | 180 | 166 | 174 | 162 |
| Turning holders | Men $n = 30$ | 34 | 33 | 37 | 38 | 41 | 43 | 46 | 43 |
| | Women $n = 18$ | 20 | 22 | 24 | 20 | 26 | 29 | 31 | 29 |
| Assembling brackets | Men $n = 40$ | 106 | 112 | 117 | 117 | 120 | 122 | 123 | 129 |
| | Women $n = 24$ | 100 | 103 | 104 | 97 | 108 | 97 | 107 | 109 |
| Assembling electrical panels | Men $n = 40$ | 2.2 | 2.7 | 2.9 | 2.8 | 2.9 | 3.0 | 3.0 | 3.0 |
| | Women $n = 24$ | 1.5 | 1.8 | 1.7 | 1.6 | 1.9 | 1.8 | 2.0 | 2.2 |
| Making wooden frames | Men $n = 40$ | 1.04 | 1.37 | 1.42 | 1.51 | 1.58 | 1.69 | 1.69 | 2.00 |
| | Women $n = 24$ | 0.38 | 0.42 | 0.50 | 0.55 | 0.69 | 0.60 | 0.74 | 0.85 |
| Embroidering cat motifs | Men $n = 40$ | 1.1 | 0.9 | 1.3 | 1.2 | 1.4 | 1.2 | 1.4 | 1.2 |
| | Women $n = 24$ | 2.2 | 1.9 | 2.3 | 2.2 | 2.2 | 2.2 | 2.0 | 2.0 |
| Packing blocks | Men $n = 40$ | 10.8 | 11.6 | 12.7 | 12.5 | 13.7 | 13.6 | 14.2 | 15.3 |
| | Women $n = 24$ | 9.4 | 11.1 | 13.3 | 11.8 | 13.9 | 12.6 | 13.8 | 13.8 |

possessed in the initial situation certain basic skills which the men had to practice. The previous familiarity with tasks of a similar kind may, finally, have contributed to the work motivation showing a tendency to decline towards the end of the working period.

In some of the tasks, it is possible to discern, alongside the training effects, also another systematic tendency which competes with these effects and which means

*Table 14.* The average quantitative performances during each of 4 consecutive days and the corresponding percentage work contribution in relation to the performance expected of the normal worker (according to the calculated MTM value).

| Working period no. | | Day 1 | | Day 2 | | Day 3 | | Day 4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | No. of products per hour | % of MTM value | No. of products per hour | % of MTM value | No. of products per hour | % of MTM value | No. of products per hour | % of MTM value |
| Pressing eyelets | Men $n = 40$ | 255 | 30 | 277 | 33 | 285 | 34 | 311 | 37 |
| | Women $n = 24$ | 228 | 27 | 252 | 30 | 275 | 33 | 280 | 33 |
| Cutting threads in nuts | Men $n = 40$ | 185 | 57 | 191 | 58 | 198 | 61 | 199 | 61 |
| | Women $n = 24$ | 151 | 46 | 162 | 50 | 173 | 53 | 168 | 51 |
| Turning holders | Men $n = 30$ | 34 | 29 | 38 | 32 | 42 | 36 | 45 | 38 |
| | Women $n = 18$ | 21 | 18 | 22 | 19 | 27 | 23 | 30 | 25 |
| Assembling brackets | Men $n = 40$ | 109 | 37 | 117 | 40 | 121 | 41 | 126 | 43 |
| | Women $n = 24$ | 101 | 34 | 100 | 34 | 102 | 35 | 108 | 37 |
| Assembling electrical panels | Men $n = 40$ | 2.4 | 19 | 2.8 | 23 | 3.0 | 24 | 3.0 | 24 |
| | Women $n = 24$ | 1.6 | 13 | 1.7 | 14 | 1.8 | 15 | 2.1 | 17 |
| Making wooden frames | Men $n = 40$ | 1.20 | 27 | 1.46 | 33 | 1.63 | 37 | 1.85 | 42 |
| | Women $n = 24$ | 0.40 | 9 | 0.52 | 12 | 0.64 | 15 | 0.80 | 18 |
| Embroidering cat motifs | Men $n = 40$ | 1.0 | 7 | 1.3 | 9 | 1.3 | 9 | 1.3 | 9 |
| | Women $n = 24$ | 2.1 | 15 | 2.2 | 16 | 2.2 | 16 | 2.0 | 15 |
| Packing blocks | Men $n = 40$ | 11.2 | 24 | 12.6 | 27 | 13.6 | 29 | 14.7 | 31 |
| | Women $n = 24$ | 10.3 | 22 | 12.5 | 26 | 13.2 | 28 | 13.8 | 29 |

that the outputs are lower during the afternoon periods than during the morning periods. This occurs only exceptionally during the first working day, which is probably connected with the fact that the training effects are then very evident.

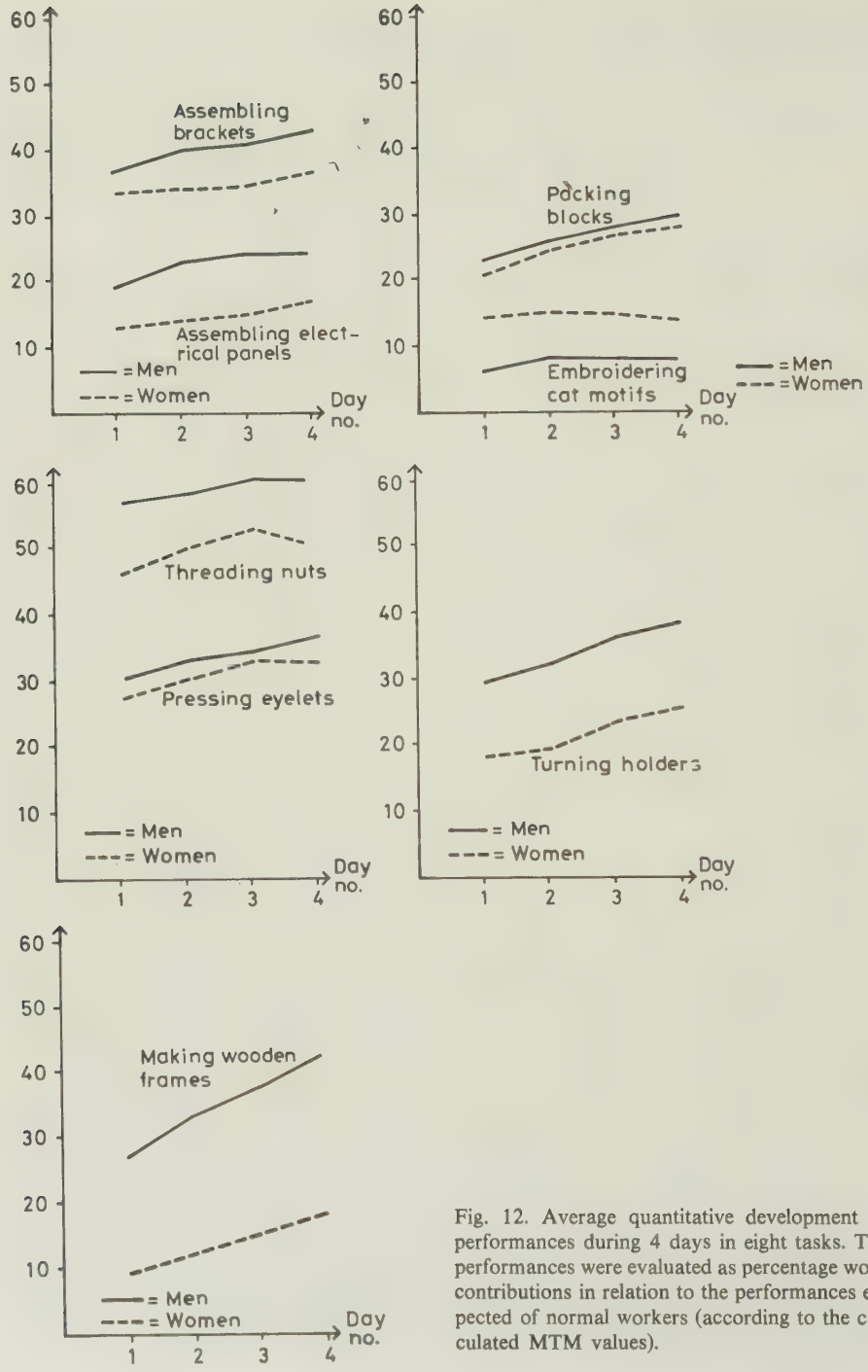



Fig. 12. Average quantitative development of performances during 4 days in eight tasks. The performances were evaluated as percentage work contributions in relation to the performances expected of normal workers (according to the calculated MTM values).

During the following 3 working days, however, this tendency appears partly sporadically in several tasks and partly as a general phenomenon in the pressing of eyelets, the cutting of threads in nuts and—as far as the men are concerned—the embroidering of cat motifs. The likeliest explanation is that this is the result of fatigue gradually setting in with time or declining motivation to work.

In order to facilitate direct comparisons between the different tasks, as regards the extent and form of the training effects, I have in Table 14 converted the average number of approved products per hour during each of the 4 working days into a percentage work contribution in relation to the anticipated performance of a normal worker given by the MTM value. From the percentages calculated in this way, I then constructed the development of performance diagrams in Fig. 12. Here I obtained a confirmation of what I have stated previously, viz. that no training effects appear among the women in the embroidering of cat motifs. Also in the male experimental group, the improvement of output in this case is extremely modest. In the cutting of threads in nuts and in the two assembly tasks, on the other hand, the training effects are clearly demonstrable. This is even more clearly the case in the pressing of eyelets and the packing of blocks. Finally, we find the strongest training effects in the making of wooden frames and the turning of holders. Here it may be recalled that it was precisely in these two tasks that we were able to note qualitative improvements.

*Measurement of quantitative performance*

As regards the form of the development of performance diagrams, it may be noted that the largest training effects appear at the beginning of the working week. Towards the end of the week, a tendency for the performances to become more stabilized may be discerned. With this in mind, I calculated measurements of the quantitative work performances in the different tasks as the numbers of approved products per hour during the last 2 working days taken together.

The question of the extent to which the level of the work performances would be changed by further training is, of course, important. Without being able to go into this question as thoroughly as I would have wished, I devoted two small special investigations to it, which I report in Appendix E. The results of these investigations indicate that training in this type of task has only a slight performance-improving effect, as soon as the worker has surmounted the initial difficulties. Though it may be expected that the performance level will rise with further training in such tasks as turning holders and making wooden frames, I think that the measurements which I calculated are to be regarded as realistic expressions of the general level of performance which the participants in the experiment may be expected to achieve after basic training in a favourable working environment. In the longer term variations in the work performances would seem to be determined more by motivation factors than by training factors.

*Table 15.* The quantitative measurement of performances, calculated as the average number of approved products per hour during the last 2 working days in eight tasks.

| Tasks | Men, $n = 40$ | | | | Women, $n = 24$ | | | | Sex difference |
|---|---|---|---|---|---|---|---|---|---|
| | Md | M | S.D | Range | Md | M | S.D. | Range | *t*-values |
| Pressing eyelets | 263.50 | 295.38 | 138.11 | 508 | 269.50 | 278.17 | 103.17 | 415 | 0.52 |
| Cutting threads in nuts | 206.00 | 198.72 | 70.50 | 299 | 174.50 | 171.08 | 60.63 | 207 | 1.57 |
| Turning holders | 42.83 | 43.55 | 21.52 | 86 | 31.50 | 29.71 | 16.63 | 60 | 2.66 |
| Assembling brackets | 113.50 | 123.72 | 50.68 | 231 | 89.50 | 104.58 | 38.37 | 130 | 1.57 |
| Assembling electrical panels | 2.95 | 3.02 | 1.82 | 7.7 | 1.45 | 1.95 | 1.25 | 4.1 | 2.50 |
| Making wooden frames | 1.55 | 1.72 | 1.24 | 4.2 | 0.45 | 0.72 | 0.64 | 2.3 | 3.62 |
| Embroidering cat motifs | 0.85 | 1.32 | 1.41 | 5.6 | 2.35 | 2.08 | 1.48 | 4.5 | –2.01 |
| Packing blocks | 13.17 | 14.15 | 8.98 | 41 | 13.50 | 13.63 | 8.55 | 35 | 0.23 |

Frequency distributions of these quantitative measurements will be found in connection with the detailed descriptions of the respective tasks in Appendix C. In addition, a summary of the median values, mean values, standard deviations and ranges of variation in men and women is given in Table 15. A scrutiny of these data shows that the measurements constructed in this way display a satisfactory ability to differentiate and give rise to distributions which are approximately normal in the great majority of cases. However, tendencies to positive skewness appear and to a marked extent in two tasks. As we already know, the task of making wooden frames proved to be almost too difficult for the women when the training period was limited to so short a period as 4 days. Likewise, the task of embroidering cat motifs involved very obvious difficulties for the men. In these two cases, therefore, the distributions were obviously positively skewed.

The variation measurements in Table 15 indicate that the inter-individual differences are remarkably large in all the tasks. If we start with the extreme values of the distributions and convert them into a percentage work contribution in relation to the normal outputs expected from the MTM analysis, we find that among the men the working capacity varies from satisfactory or almost satisfactory productivity to complete incapacity to put up approved performances. Among the women, the corresponding variation is certainly somewhat less but is still very considerable. This confirms what I already thought that I had found in connection with the analysis of the data derived from the inquiry sent to all the sheltered workshops in Sweden in 1967 (Palmér 1968*a*), viz. that we have to allow for a very wide total spread of the working capacity among the mentally retarded.

*Estimate of reliability*

In order to estimate the reliability of the measurements of quantitative capacity just presented, I have correlated the performance measurements worked out in an analogous manner during each of the last 2 working days and have reported the results in Table 16. The values, which have been corrected in the usual way by the Spearman-Brown formula, vary between .94 and .98. It is accordingly obvious that the reliability of these criterion measurements may reasonably be said to be exceptionally satisfactory.

*Sex differences and intercorrelations*

If we return to Table 15 and scrutinize it regarding the occurrence of sex differences, we may observe that the women performaed better than the men in embroidering cat motifs, while the state of affairs was reversed in the other tasks. This is the pattern of differences which we have previously encountered several times in Tables 13 and 14, with the successive measurements of performance reported there.

*Table 16.* The product-moment coefficients of correlation between quantitative performances during each of the last 2 working days. Total experimental group, $n = 64$.

| Task | $r$ | $r_c$ |
|---|---|---|
| Pressing eyelets | .91 | .95 |
| Cutting threads in nuts | .95 | .97 |
| Turning holders | .89 | .94 |
| Assembling brackets | .94 | .97 |
| Assembling electrical panels | .96 | .98 |
| Making wooden frames | .96 | .98 |
| Embroidering cat motifs | .90 | .95 |
| Packing blocks | .96 | .98 |

One may get an immediate comparison of the magnitudes of the differences in the various tasks by studying the diagrams in Fig. 12, which represent the performances during each of the four working days as percentages of the calculated MTM values. In order to further facilitate such comparisons, I have calculated and inserted the $t$-values of the differences in Table 15. Whichever of these two sources of information we use, we find that the sex differences in the packing of blocks and the pressing of eyelets are rather slight, while they are far more marked in the making of wooden frames and the turning of holders. In the four remaining tasks, the differences are of moderate size. They reach a statistically significant level in both the making of wooden frames and the turning of holders, as well as the assembling of electrical panels and the embroidering of cat motifs.

It is also interesting to study the intercorrelations between the performances in the eight tasks. I have reported these separately for men and women in Table 17. In the male group, these intercorrelations are positive and comparatively high throughout. They vary between .50 (turning holders/embroidering cat motifs) and .86 (assembling electrical panels/making wooden frames). Thus, there is a fairly strong probability that an individual who is successful in one of the eight tasks will also be successful in the others. Likewise, all the intercorrelations are positive in the female group. The majority of them are, however, lower than the corresponding values in the male group and in some cases do not reach a statistically significant level. Cutting threads in nuts in particular but also assembling brackets show low correlations with the other tasks.

In the correlation matrix, I have also included the prognosis-group and the percentage-group criteria. If we consider the correlations between these and the performances in the eight tasks, another and perhaps more remarkable difference between men and women appears. Thus, in the men, these correlations—with one exception—are high throughout. The exception is the assembly of brackets. The correlations between this task and the two rating criteria are

*Table 17.* Correlation matrix for performance in eight tasks and the prognosis-group and percentage-group criteria among men ($n$ = 40) (below the diagonal) and women ($n$ = 24) (above the diagonal).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Pressing eyelets | | .41 | .53 | .63 | .68 | .61 | .59 | .55 | .39 | .36 |
| 2. Cutting threads in nuts | .62 | | .16 | .51 | .26 | .24 | .48 | .22 | .19 | .36 |
| 3. Turning holders | .60 | .64 | | .33 | .79 | .57 | .57 | .57 | .23 | .11 |
| 4. Assembling brackets | .54 | .69 | .59 | | .33 | .27 | .63 | .43 | .20 | .40 |
| 5. Assembling electrical panels | .60 | .67 | .59 | .57 | | .83 | .46 | .65 | .47 | .20 |
| 6. Making wooden frames | .57 | .59 | .66 | .54 | .86 | | .39 | .60 | .54 | .29 |
| 7. Embroidering cat motifs | .70 | .64 | .50 | .52 | .63 | .64 | | .43 | .29 | .34 |
| 8. Packing blocks | .56 | .75 | .66 | .55 | .70 | .72 | .68 | | .42 | .53 |
| 9. Prognosis group | .67 | .58 | .57 | .34 | .67 | .61 | .67 | .58 | | .55 |
| 10. Percentage group | .69 | .72 | .64 | .49 | .75 | .73 | .73 | .80 | .75 | |

.34 and .49 respectively; they cannot be described as high, but nevertheless they also reach a clearly significant order of magnitude. Furthermore, we may note that the correlations between the performances and the percentage-group criterion are in all cases somewhat higher than those referring to the prognosis-group criterion. This supports my previous assumption that the percentage-group criterion is a purer measure of performance than the prognosis-group criterion (cf. p. 53 above). If we then consider the corresponding conditions in the female group, we find that the correlations between the performance and the two rating criteria are in most cases low. Thus, the prognosis-group criterion is correlated, in a statistically significant sense, only with the assembling of electrical panels, the making of wooden frames, and the packing of blocks. The difference demonstrated in the male group between the two rating criteria with consistently higher correlation coefficients between the percentage-group criterion and the performances has here been replaced by an irregular pattern. These circumstances may mean that the judgement processes which resulted in the prognosis-group and percentage-group criteria followed a different course, depending on whether the judgements referred to men or to women. It is not out

of the question that the different expectations of, or attitudes to, the behaviour of men and women in their working life, which are prevalent in our society, may have perceptibly influenced the raters and helped to form their frame of reference.

However, it is not possible to state with certainty that the correlation differences reported are invariably dependent on sex. On the contrary, it is probable that they depend—perhaps to a considerable degree—on the lower variability in the female group. As I pointed out above (cf. p. 65), the possibilities of making comparisons between men and women are more limited, since in this part of the investigation I was compelled to reduce the total experimental group by seven men and nine women. It is true that this did not alter the average values in the variables of the prognosis group and the percentage group, either in the male or the female groups, but the already low variability in the female group became even lower, relatively speaking. This fact justifies me in expressing an opinion on existing sex differences in only very guarded terms as regards intercorrelations. It also gives me reason to devote most of my attention to the male group in the subsequent treatment of the data. At the same time, however, I shall study the female group and also the total experimental group, for purposes of comparison.

*Table 18.* Correlation matrix for quantitative measurements of performance in eight tasks and the prognosis-group and percentage-group criteria. The total experimental group ($n$ = 64).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Pressing eyelets | | | | | | | | | | |
| 2. Cutting threads in nuts | .56 | | | | | | | | | |
| 3. Turning holders | .57 | .53 | | | | | | | | |
| 4. Assembling brackets | .56 | .65 | .55 | | | | | | | |
| 5. Assembling electrical panels | .61 | .59 | .68 | .54 | | | | | | |
| 6. Making wooden frames | .55 | .53 | .66 | .55 | .86 | | | | | |
| 7. Embroidering cat motifs | .62 | .51 | .39 | .47 | .45 | .39 | | | | |
| 8. Packing blocks | .56 | .57 | .61 | .51 | .66 | .62 | .56 | | | |
| 9. Prognosis group | .60 | .50 | .52 | .34 | .65 | .62 | .45 | .52 | | |
| 10. Percentage group | .60 | .63 | .52 | .49 | .62 | .64 | .52 | .70 | .71 | |

I have accordingly reproduced in Table 18 the same correlation matrix, calculated on the total experimental group. Here we find that all the coefficients are clearly significant at the 1 % level. However, they are, as expected, generally lower than the corresponding values in the male group. The reduction chiefly affects the embroidering of cat motifs and then the cutting of threads in nuts. This would seem to indicate that these two tasks may be said to be in some sense more sex-specific than the rest. Sewing has long been regarded as a typically female task. As I have previously pointed out, this means that girls have more practice in sewing than boys, as they grow up. This not only leads to their having greater skill but also influences their emotional attitude. In the present investigation, it was especially obvious that, without exception, the women felt more confident in the sewing task than in the other tasks. To a certain extent, the opposite was true, as regarded all the three tasks involving machine work and the making of wooden frames. However, there were signs that cutting threads in nuts was regarded by the women as more foreign to them than the other tasks. Thus, when, in a special attitude test after the completion of the testing programme, they had to comment on which task they would prefer to continue with, they placed the embroidering of cat motifs first, while cutting threads in nuts came at the bottom of the list. I received a definite impression that the deprecating attitude to the latter task was to some extent connected with the fact that it was oily and required the wearing of a protective apron.

*Dimensions of work performance*

The above examination of the intercorrelations between the tasks raises the interesting question of whether it is possible, by a study of this kind, to demonstrate any dimensionality in the work performance of the mentally retarded. I consciously endeavoured to secure variety in the tasks—not only as regarded the degree of difficulty but also as regarded the basic material, the length of the work cycle, the requirements for precision and co-ordination of movements etc.—and I had certain expectations in this direction. I fully realize that the investigation—*inter alia,* on account of its quantitative limitations—does not shed any definitive light on this problem. Nevertheless, I was unwilling to omit to carry out, with a markedly exploratory object, factor analyses of the correlations between the performances in the eight tasks. In doing so, I applied Jöreskog's J method (Jöreskog 1963) and used the FAIM programme. The results are shown in Table 19, the different parts of which give the eigenvalues and the significance tests of the number of factors, the unrotated and rotated factor matrices and the estimated communalities.

In order to estimate the number of factors required to explain the common variance, we may first study the distribution of the eigenvalues. In this connection, we should remember that the J method presupposes that $k$ (the number of

factors) is so large that $T_k < 1$. This means that $k \geqslant 2$. The significance test supports the hypothesis $k = 2$ in the female group but not in the male group. Nevertheless it seems reasonable to assume two factors in each group.

If we examine the rotated factor matrices—seeking first the common denominator among men and women—we find that the first factor had its greatest loadings in both men and women in the assembly of electrical panels and the making of wooden frames. The characteristic feature of these two tasks is that they consist of several subsidiary operations and also require the use of several

*Table 19.* Factor analysis of quantitative measurements of performance in eight tasks.

(*a*) Eigenvalues and significance tests of the number of factors.

| Men, $n = 40$ | | | Women, $n = 24$ | | |
|---|---|---|---|---|---|
| Eigenvalues | Diff. | | Eigenvalues | Diff. | |
| 19.02 | 16.70 | | 16.94 | 13.31 | |
| 2.32 | 0.80 | | 3.63 | 1.76 | |
| 1.51 | 0.22 | | 1.87 | 0.74 | |
| 1.30 | 0.17 | | 1.13 | 0.16 | |
| 1.12 | 0.23 | | 0.98 | 0.11 | |
| 0.89 | 0.36 | | 0.86 | 0.26 | |
| 0.54 | 0.08 | | 0.60 | 0.19 | |
| 0.45 | 0.0 | | 0.42 | 0.0 | |
| $T_K$ | $D_K$ | $U_K$ | $T_K$ | $D_K$ | $U_K$ |
| 1.16 | 27 | 36.24 | 1.36 | 27 | 47.85 |
| 0.97 | 20 | 18.91 | 0.98 | 20 | 16.31 |
| 0.86 | 14 | 14.39 | 0.80 | 14 | 6.28 |
| 0.75 | 9 | 10.46 | 0.71 | 9 | 4.51 |

$T_K$ = Jöreskogs $T_K$; $D_K$ = Degrees of freedom; $U_K = \chi^2$

(*b*) Factor matrix, men. $n = 40$.

| Task | Unrotated factor loadings | | Rotated factor loadings | | $h^2$ |
|---|---|---|---|---|---|
| | 1 | 2 | 1 | 2 | |
| 1. Cutting | .82 | −.30 | .37 | −.79 | .76 |
| 2. Embroidering | .77 | −.12 | .46 | −.63 | .61 |
| 3. Electrical panels | .88 | .24 | .79 | −.45 | .82 |
| 4. Packing | .84 | −.08 | .54 | −.65 | .71 |
| 5. Wooden frames | .87 | .29 | .82 | −.41 | .85 |
| 6. Brackets | .71 | −.20 | .36 | −.64 | .55 |
| 7. Pressing | .73 | −.19 | .38 | −.65 | .57 |
| 8. Turning | .74 | −.13 | .43 | −.62 | .57 |
| Total contribution of variance | 5.09 | 0.35 | 2.40 | 3.04 | 5.44 |

(*c*) Factor matrix, women. $n = 24$.

| Task | Unrotated factor loadings | | Rotated factor loadings | | $h^2$ |
|---|---|---|---|---|---|
| | 1 | 2 | 1 | 2 | |
| 1. Cutting | .37 | .48 | .12 | .59 | .37 |
| 2. Embroidering | .63 | .47 | .36 | .70 | .62 |
| 3. Electrical panels | .92 | −.21 | .92 | .21 | .89 |
| 4. Packing | .71 | .02 | .63 | .33 | .51 |
| 5. Wooden frames | .82 | −.22 | .83 | .17 | .71 |
| 6. Brackets | .51 | .63 | .18 | .79 | .66 |
| 7. Pressing | .78 | .28 | .58 | .59 | .68 |
| 8. Turning | .81 | −.10 | .77 | .26 | .66 |
| Total contribution of variance | 4.08 | 1.03 | 3.04 | 2.07 | 5.11 |

different kinds of tool. The second factor has its greatest loadings in both men and women in the cutting of threads in nuts, the assembling of brackets and the embroidering of cat motifs. The first two of these tasks have markedly short cycles and—as we have already seen—do not call for work of high quality. The fact that the embroidering of cat motifs has heavy loadings in this factor may at first seem surprising. This is a task with a comparatively long cycle and it gave the men in particular considerable difficulties. However, it does not require the use of several different kinds of tool and is fairly uniform, apart from the machining operation. The remaining three tasks—packing, pressing and turning—present a more obscure picture which shows certain differences as between the men and the women. Thus, in the male group all three have higher loadings in the second than in the first factor. In the female group, on the other hand, turning holders and packing blocks are more or less definitely connected with the first factor, whereas pressing eyelets lies somewhere in between. In this connection, we may also note that in the female group it is the first factor which is responsible for the major contributions of variance, while the opposite state of affairs exists in the male group. In themselves, such differences need not surprise us. They have, of course, a varying causal backgroud but may, amongst other things, be connected with differences between men and women as regards acquired experience during childhood and adolescence. Here it may be recalled that Fleishman and Hempel (1954) have demonstrated how the factor structure in psychomotor functions may be dependent on the amount of training which the experimental subjects have received.

It would seem to be unjustifiable in any circumstances to draw, solely from this study, any definite conclusions concerning the factor structure of the capacity for work of the mentally retarded. However, we have got a good basis for the hypothesis that in pure "factor-structure-seeking studies" (Dingman and Meyer 1966, Kääriäinen 1970) one must be able to distinguish two factors, one of which

represents the ability to carry out successfully tasks which consist of several subsidiary operations and which presuppose the use of several different hand tools, while the other involves instead the ability to cope with tasks with relatively short work cycles, which do not include the handling of tools at all or only include such handling to a small extent.

*Composite measurement of work performance*

The measurements of quantitative performances in eight tasks reported in the previous section are each basic criterion variables in the present investigation and will be dealt with as such in the following studies. However, there are at the same time grounds for regarding the eight tasks as so many sub-tests in a test battery. In the factor analysis of the measurements of quantitative performance just presented, there were signs that at least two working capacity factors can be distinguished. On the basis of this result, I may be justified in classifying the tasks in two groups, each designed to measure its own factor. As I have already pointed out, however, there is reason to regard the results of the factor analysis as tentative. As a matter of fact, some tasks also display equally heavy loadings in both the factors. For these reasons, I find it best to treat the test battery as an omnibus-test scale, in which the different sub-performances are more or less closely correlated with each other, and in which the average of them may constitute a measurement of general working capacity.

In order to obtain such a composite measurement, I have converted the individual raw scores in the eight tasks into the corresponding $z$ values and then calculated the averages of these values. Having avoided the disadvantages of standard scores in the usual way, by expressing them on a scale, with 50 as the mean value and 10 as the standard deviation, we may illustrate their distributions as shown in Fig. 13.

As we know, transformations of this kind do not entail any change in the forms of distribution. We have already noted that the raw-score distributions were approximately normal in most cases. In so far as deviations from the normal form of distribution occurred, they were tendencies to positive skewness. In two tasks, such tendencies appeared to a marked degree. This applied to the embroidering of cat motifs, which gave the men obvious difficulties. The same state of affairs was found—most markedly amongst the women—as regarded the making of wooden frames. It was these circumstances which led to the distributions of the average $z$ values in Fig. 13 also displaying a tendency to positive skewness, a tendency which, however, is only moderately marked. As might have been expected, the mean value is higher in the male group than in the female group. However, the difference is not statistically significant.

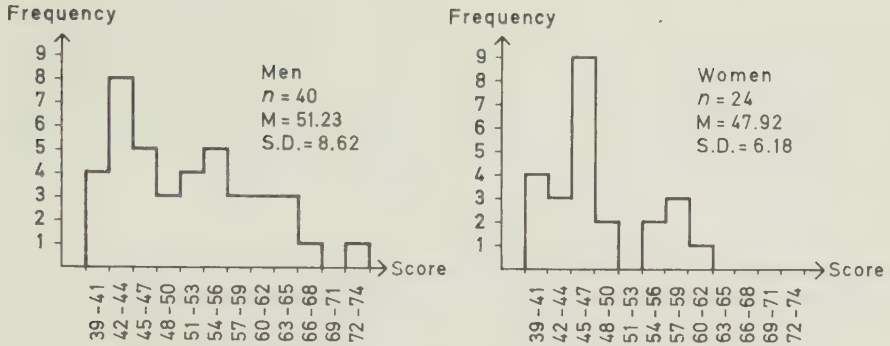



Fig. 13. Distributions of average quantitative performances (average of *z*-values in 8 tasks) in mentally retarded men and women.

## 3. Observations of Behaviour and Adjustment during the Testing Programme

In the discussion in the introductory section on the problems concerning criteria, I rejected the idea of designing a criterion variable which was wide enough to cover the whole complex meaning of the concept of social adjustment. Instead, I considered that there were over-riding reasons for restricting myself to the observation and measurement of work performances and adjustment to work. In these circumstances, the main attention will generally be concentrated on the performance aspect. One reason for this is that in this field there are facilities for making more reliable observations and measurements. However, it may also be argued that the performance aspect is the most important from the point of view of habilitation. As I have already pointed out in another connection, it is his quantitative performance that will chiefly decide the individual's prospects of success on the labour market. Indirectly, it is consequently also one of the most important preconditions for both the experience of personal satisfaction and for social adjustment in the wider sense.

However, this does not by any means imply that I recommend that in prediction investigations of this kind one should entirely disregard questions concerning job satisfaction and adjustment to work. These aspects of behaviour are obviously of extremely great interest from both the practical and the theoretical points of

*Table 20.* Estimates of reliability (alphas) on a scale which measures adjustment to work. in eight tasks.

| Task | Pressing eyelets | Cutting threads in nuts | Turning holders | Assembling brackets | Assembling electrical panels | Making wooden frames | Embroidering cat motifs | Packing blocks |
|---|---|---|---|---|---|---|---|---|
| Men $n = 40$ | .87 | .88 | .93 | .94 | .90 | .89 | .92 | .96 |
| Women $n = 24$ | .94 | .97 | .96 | .95 | .94 | .92 | .94 | .95 |

view. I also made use of the opportunities of studying them closely, by making systematic observations of the behaviour in the eight tasks.

As an instrument in these observations, I used certain parts of the general observation form of the check-list type, which I had previously used to develop prediction variables (cf. above p. 47ff.). Seventy-nine of the original 155 items were included in this version. A research technician, a psychologist and a supervisor acted as observers. The first two observed all the participants in the testing programme in each of the eight tasks. The supervisor, on the other hand, changed from task to task. The data obtained in this way were analysed by the same technique as was applied in connection with the design of the predictor variables. I was then able to make a scale which consists of 39 items and whose contents cover various aspects of motivation and emotional adjustment in work situations (Holm and Palmér 1972). As might have been expected from previous

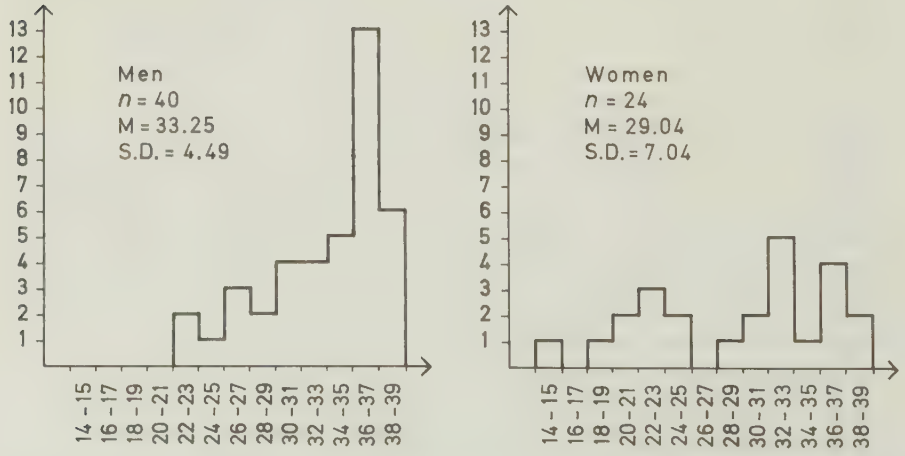



Fig. 14. Distributions of average ratings of adjustment to work, I (using the check-list form) among men and women.

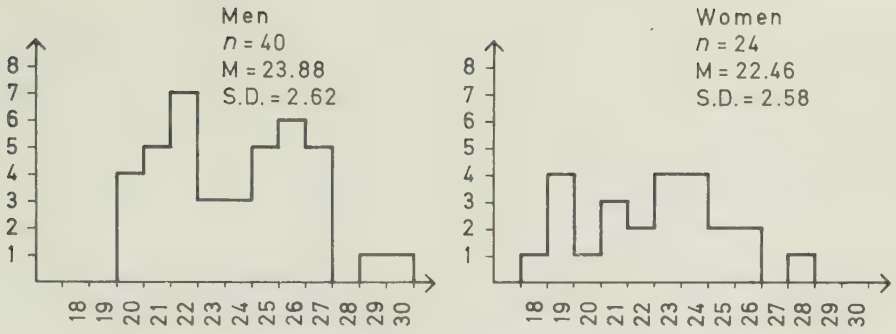



Fig. 15. Distributions of average ratings of adjustment to work, II (using the five-degree rating schedule) among men and women.

experience, the distributions were negatively skewed in this case also, while the calculated measurements of reliability (coefficient alpha) were large. In Table 20, it will be found that they vary between .87 and .97.

Since in this prediction study I was primarily interested in obtaining a composite measure of the degree of adjustment to work shown in different work situations, I calculated for each individual the average of the scale scores worked out in the eight tasks. The values obtained in this way are distributed as shown in Fig. 14. It may be noted that the men in particular display a marked tendency to a negatively skewed distribution. The mean value is higher in the male group. The difference is statistically significant at the 2 % level.

In parallel with these observations connected with the shortened version of the check-list form, I also observed the participants in the testing programme in all the tasks, using the five-degree rating schedule (cf. above, p. 51 and Appendix B:2). As in the design of the predictor variables, it was possible here to distinguish three sub-scales, one of which may be said to measure adjustment to work, since it contains items referring to persistence, interest in the work, orderliness, punctuality, general conduct and ability to co-operate. The estimates of reliability (coefficient alpha) vary on this scale from .81 in the making of wooden frames to .89 in the assembling of electrical panels and the embroidering of cat motifs. In an endeavour to obtain a composite measure, I also calculated in this case the averages of individual scale scores in all the tasks, which resulted in values distributed as shown in Fig. 15. These distributions may be described as being approximately normal. Here also the men display a higher value than the women. The difference is significant at the 5 % level.

Finally I tried in a third way to secure a measure of the participants' adjustment to work. I instructed the supervisors, from the beginning of the investigation, to notice and record any interruptions that might occur in the work.

However, there was no question of carrying out a minute check. In every workplace, it is natural and inevitable that people should take short breaks, for example, to go to the toilet, to have a smoke or to merely open the window and breathe some fresh air. This was, of course, also tolerated during the testing programme. But if these legitimate breaks extended over five minutes and the worker did not then return to his task, in spite of being told to do so, the duration of the interruption was recorded on a special form.

Such interruptions occurred relatively seldom, in consequence of the high level of motivation which characterized most of the workers. It is therefore impossible to make a reliable comparison between different tasks from this point of view. However, I have calculated the total time taken for these interruptions for all the tasks and have given it in minutes. The values obtained in this way are distributed in the way shown in Fig. 16. We are constrained to state that they differ very markedly from normal distributions, in that they are extensive and also skewed. However, they are the result of basic observations within a field of behaviour which is both important and difficult of access, and they display—as we shall see later on—consistent correlations with other criterion variables. The women's values show a longer average period of stoppage and a wider range than the corresponding values for the men. The difference is statistically significant at the 5 % level. Thus, we may state that the men display a greater adjustment to work than the women, whichever of the three measures just reported we use.

Before I conclude my account of the observations made during the testing programme, I must mention that I also calculated—in order to shed more light on the structure of the criterion variables—three composite measures referring to aspects of performance in the behaviour at work. Thus, in connection with the statistical treatment of the check-list form, it proved to be possible to form a scale of 16 items, whose contents measure the proficiency shown, chiefly from the qualitative point of view. After calculating the average scale values for the eight tasks, I found that they are distributed as shown in the histograms in Fig. 17. The distributions are negatively skewed, particularly in the male group. The mean values in men and women do not differ significantly from each other, but the range is obviously greater among the men.

I also calculated composite measures in the two remaining sub-scales of the five-degree rating schedule, which measure proficiency and quantitative performance. The average values of the former sub-scale are distributed as shown in the histograms in Fig. 18. As in the previous measurement, the mean value in the male group is somewhat higher than in the female group, though the difference does not achieve statistical significance.

As I have already shown, the rating of the quantitative performance was carried out in connection with a single item with five scale steps. In working out the average measurement for the eight tasks, I multiplied the quotient by 10 and

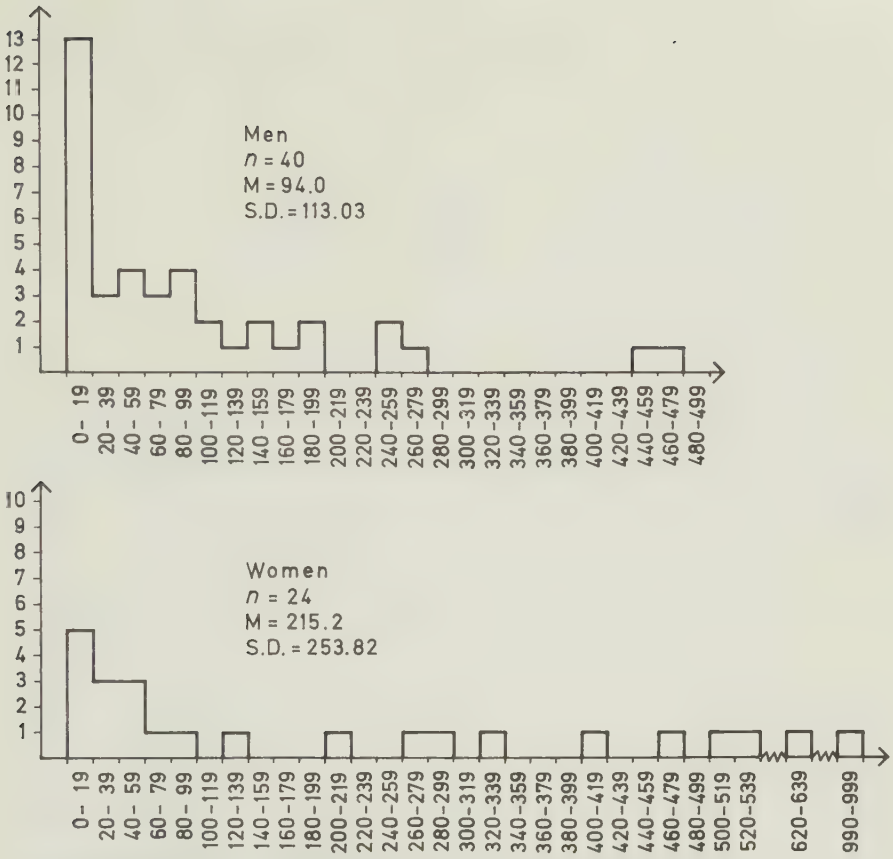



Fig. 16. Distributions of total stoppage time in minutes among men and women.

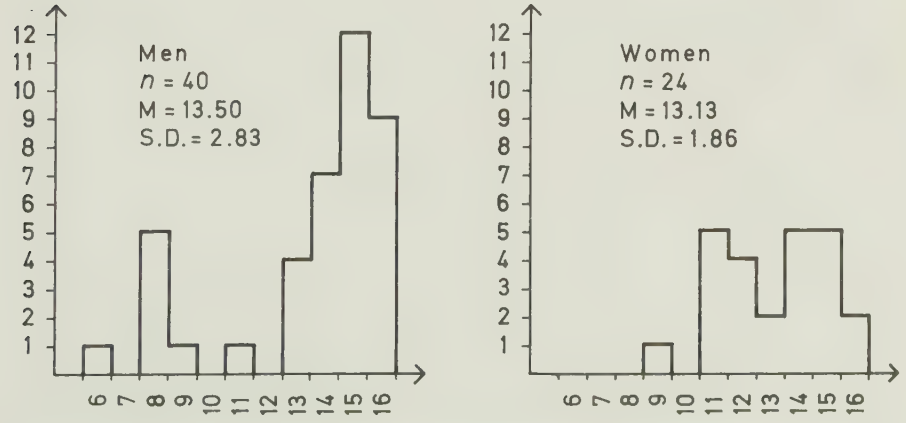



Fig. 17. Distributions of average ratings of work proficiency, I (using the check-list form).

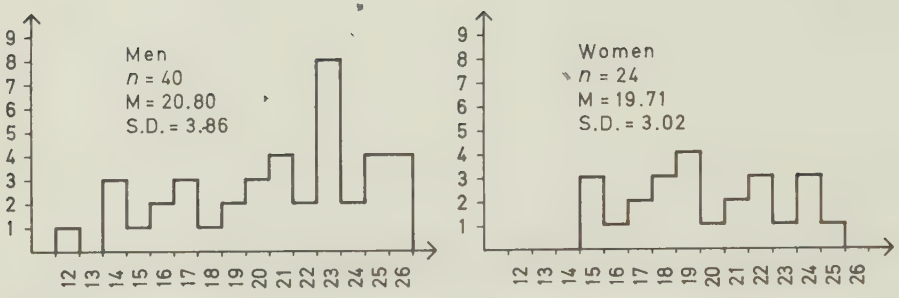



Fig. 18. Distributions of average ratings of work proficiency, II (using the five-degree rating schedule).

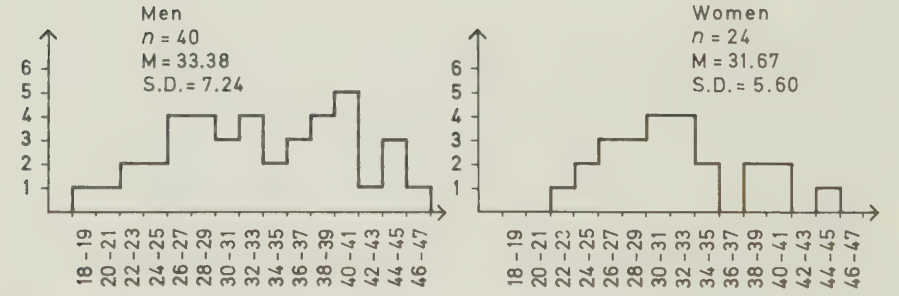



Fig. 19. Distributions of average ratings of quantitative performances in eight tasks.

obtained the distribution shown in Fig. 19. The men display a somewhat higher value than the women, but the difference is not statistically significant.

## 4. Summary Analysis of Criterion Variables

In the foregoing pages, I have described the genesis of a large number of criterion variables. They may be assigned to three groups, as follow:

1. Rating criterion variables based on general observations of behaviour at work in a sheltered workshop.
   - 1:1. The prognosis-group criterion
   - 1:2. The percentage-group criterion
   - 1:3. The wages-group criterion
2. Quantitatively determined performances in the eight tasks.
   - 2:1. Pressing eyelets
   - 2:2. Cutting threads in nuts
   - 2:3. Turning holders
   - 2:4. Assembling brackets
   - 2:5. Assembling electrical panels
   - 2:6. Making wooden frames
   - 2:7. Embroidering cat motifs
   - 2:8. Packing blocks
3. Composite criterion measures based on measurements of performance and systematic observations of behaviour in the eight tasks.
   - 3:1. Average quantitative performance
   - 3:2. Average qualitative performance
   - 3:3. Average rating of adjustment to work, I (using the check-list form)
   - 3:4. Average rating of adjustment to work, II (using the five-degree rating schedule)
   - 3:5. Total stoppage time
   - 3:6 Average rating of work proficiency in the tasks, I (using the check-list form)
   - 3:7 Average rating of work proficiency in the tasks, II (using the five-degree rating schedule)
   - 3:8 Average rating of quantitive performance in the tasks.

Let us first turn our attention to the rating criteria and bear in mind that these criteria came into existence before the testing programme which included the eight specially selected tasks was carried out. Thus, the observations on which the ratings are based were made in connection with the regular workshop operations. It should also be borne in mind that, in consequence of a gradual change in the practice with regard to wages, I was not in a position to procure any values in the wages-group criterion for the whole experimental group (cf. above, pages 55—57). This fact justifies me in omitting this variable, on the whole, from the subsequent studies. This is, of course, regrettable from some points of view, but the loss is probably not very great, in view of the fact that the wages criterion—as Table 8 showed—displayed substantial or high correlations with both the prognosis-group criterion (.76 among men and .68 among women) and the percentage-group criterion (.70 among men and .80 among women). As regards the two latter criteria, I have previously reported relatively high correlations between them also (in the complete experimental group, .74 among the men and .61 among the women). However, there is hardly reason, on that account, for regarding them as identical in content or for letting them form a single criterion variable on that basis by means of score transformations.

In connection with my account of how the ratings were carried out, I assumed that the percentage-group criterion was a more delimited and more homogeneous measurement of quantitative performance than the prognosis-group criterion. While in the former variable the raters had only to express their opinions as to the anticipated percentage contribution of work, their task in the latter variable involved classifying the workers with respect to their suitability for placement in various institutions or jobs. It is likely that, in carrying out the latter task—with the favourable and hopeful attitude which they took to the workers—they were apt to adopt a longer perspective in time and thereby to assume that actual deficiencies could be compensated by further training or by the satisfactory functional capacity for social life already shown. This may explain why the marked positive skewness in the distribution of the percentage-group criterion is not to be found in the prognosis-group variable.

I felt I had found some support for the reported assumption of a difference in content between the two criteria when I examined the correlations between them and the quantitative performances in the eight tasks (cf. above, page 79). Moreover the same trend is seen in certain correlations between, on the one hand, the prognosis-group and percentage-group variables and, on the other, the composite criteria based on measurements of performance and observations of behaviour in the eight tasks. These correlations are shown in Table 21.

Before I turn my attention to the correlations here quoted, however, I shall examine the table regarding the intercorrelations between the composite criteria themselves. Criteria 3:1, "Average quantitative performance," 3:2, "Average

*Table 21.* Correlation matrix for two rating criterion variables and eight composite criterion variables among men ($n$=40), among women ($n$=24) and in the total group ($n$=64).

| Criterion variables | | 1:1 | 1:2 | 3:1 | 3:2 | 3:3 | 3:4 | 3:5 | 3:7 | 3:8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1:1 Prognosis-group criterion | | | | | | | | | | |
| 1:2 Percentage-group criterion | Men | .75 | | | | | | | | |
| | Women | .55 | | | | | | | | |
| | Total | .71 | | | | | | | | |
| 3:1 Average quantitative performance | Men | .71 | .84 | | | | | | | |
| | Women | .46 | .44 | | | | | | | |
| | Total | .66 | .75 | | | | | | | |
| 3:2 Average qualitative performance | Men | −.68 | −.67 | −.71 | | | | | | |
| | Women | −.42 | −.17 | −.66 | | | | | | |
| | Total | −.60 | −.52 | −.69 | | | | | | |
| 3:3 Average rating of adjustment to work, I | Men | −.19 | −.39 | −.48 | .20 | | | | | |
| | Women | −.16 | −.36 | −.29 | .00 | | | | | |
| | Total | −.24 | −.40 | −.42 | .14 | | | | | |
| 3:4 Average rating of adjustment to work, II | Men | .57 | .72 | .74 | −.56 | −.83 | | | | |
| | Women | .30 | .51 | .41 | −.15 | −.88 | | | | |
| | Total | .51 | .67 | .66 | −.43 | −.84 | | | | |
| 3:5 Total stoppage time | Men | −.11 | −.33 | −.29 | .22 | .46 | −.45 | | | |
| | Women | −.42 | −.33 | −.22 | −.04 | .59 | −.50 | | | |
| | Total | −.29 | −.33 | −.27 | .10 | .59 | −.49 | | | |
| 3:6 Average rating of work proficiency, I | Men | −.70 | −.64 | −.70 | .86 | .10 | −.48 | .15 | | |
| | Women | −.45 | −.33 | −.68 | .76 | .31 | −.48 | −.01 | | |
| | Total | −.63 | −.56 | −.69 | .82 | .18 | −.47 | .08 | | |
| 3:7 Average rating of work proficiency, II | Men | .83 | .78 | .84 | −.84 | −.22 | .59 | .15 | −.91 | |
| | Women | .63 | .55 | .70 | −.76 | −.20 | .46 | −.05 | −.88 | |
| | Total | .78 | .72 | .80 | −.82 | −.24 | .56 | −.11 | −.90 | |
| 3:8 Average rating of quantitative performance | Men | .74 | .80 | .93 | −.66 | −.49 | .74 | −.23 | −.74 | .86 |
| | Women | .46 | .57 | .85 | −.43 | −.35 | .52 | −.27 | −.63 | .68 |
| | Total | .67 | .74 | .91 | −.59 | −.42 | .67 | −.25 | −.71 | .82 |

Note. The variables 3:2, 3:3 and 3:6 were inverted before the calculations were made, which explains why their correlations with the other variables are negative (apart from Total stoppage time).

qualitative performance" and 3:5, "Total stoppage time", occupy special positions, in that they result from performed measurements and are thus but slightly influenced by subjective judgements. I can here use them in their capacity of objective measurements to give validity to the other variables.

It follows from what I said earlier that I regard the criterion variable 3:1, "Average quantitative performance", as the chief criterion in the part of the investigation which includes the testing programme. The complexity of this varia-

ble emerges from the fact that in the total group it displays significant correlations with all other subsidiary criteria. It correlates most closely—as could be foreseen—with 3:8, "Average rating of quantitative performance", and thereafter with 3:7, 3:6 and 3:2, or the variables which may be assumed to measure performance mainly from the qualitative aspect. I may also note the marked correlation with 3:4, "Average rating of adjustment to work, II", which is, however, considerably stronger among the men than among the women. Finally, there is a weak correlation between the main criterion and the two remaining variables of adjustment. These values in the female group are not statistically significant. This is also the case in the male group for the "Total stoppage time" variable.

I have already had reason to note that the criterion variable 3:2, "Average qualitative performance", displays such a comparatively high correlation with the main criterion in connection with my discussion of quantity and quality. Here I can add that the correlation found is by no means surprising if one remembers that there is a standard of quality inherent in the quantitative measurements of performance. As I stated above, these were calculated throughout as the number of *approved* products per unit of time. However, in this context I wish to direct attention particularly to the fact that variable 3:2 correlates most closely with variables 3:6 and 3:7. This confirms that the latter two rating variables actually measure performance primarily from the point of view of quality. Moreover their intercorrelation is as high as .90 in the total group. There are but slight differences between men and women in this respect.

The criterion variable 3:5, "Total stoppage time", is, as I have just noted, only weakly correlated with the main criterion and displays no correlation with the quality measures just mentioned. On the other hand, it co-varies—although only to a moderate extent—with the variables 3:3 and 3:4, which are ratings of adjustment to work. Their intercorrelation is as high as .84. There are no marked differences between men and women in these respects either.

It is now time to return to the correlations, given in Table 21, between the composite criterion measures and the prognosis-group and percentage-group variables respectively. We then find that the prognosis-group criterion displays higher correlations than the percentage-group criterion with all three variables which comprise measures of work performance mainly from the qualitative point of view, namely, 3:2, 3:6 and 3:7. On the other hand, the percentage-group criterion tends to be closely correlated with the two variables which measure primarily the quantitative performance, namely, 3:1 and 3:8. Thus, I am of the opinion that the correlations shown here indicate that the prognosis-group variable takes more account of qualitative skill and thereby of capacity for development or potential ability on the part of the individual worker, while the percentage-group variable, on the other hand, reflects to a greater extent the actual performance from the quantitative point of view. Thus, there is reason to keep them separate in the further treatments.

*Table 22.* Factor matrix for 10 criterion variables in the total group ($n$=64).

| Criterion variables | | Unrotated factor loadings | | | Rotated factor loadings | | | $h^2$ |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 1 | 2 | 3 | |
| 1:1 | Prognosis-group criterion | .77 | .12 | .01 | .63 | −.18 | .42 | .61 |
| 1:2 | Percentage-group criterion | .80 | −.11 | .15 | .49 | −.36 | .56 | .68 |
| 3:1 | Average quantitative performance | .91 | −.03 | .24 | .56 | −.31 | .69 | .88 |
| 3:2 | Average qualitative performance | −.77 | −.33 | .25 | −.85 | .06 | −.20 | .77 |
| 3:3 | Average rating of adjustment to work, I | −.47 | .79 | .16 | −.07 | .92 | −.13 | .86 |
| 3:4 | Average rating of adjustment to work, II | .75 | −.56 | −.13 | .37 | −.81 | .30 | .89 |
| 3:5 | Total stoppage time | −.29 | .53 | .07 | −.01 | .60 | −.10 | .37 |
| 3:6 | Average rating of work proficiency, I | −.84 | −.33 | .22 | −.89 | .08 | −.26 | .86 |
| 3:7 | Average rating of work proficiency, II | .93 | .27 | −.08 | .86 | −.13 | .43 | .94 |
| 3:8 | Average rating of quantitative performance | .91 | −.04 | .26 | .54 | −.31 | .71 | .90 |
| Total contribution of variance | | 5.92 | 1.53 | .31 | 3.64 | 2.24 | 1.88 | 7.76 |

Finally, in order to see whether this picture of the content of the criterion variables which I secured by studying the intercorrelations in Table 21 will also emerge, and perhaps take on greater clarity, by means of factor analysis, I carried out such an analysis according to the Jöreskog J method (Jöreskog 1963), using the FAIM programme. I felt justified in using this procedure for so limited a purpose, even though some of the distributions are rather seriously skewed and the groups are comparatively small. Bearing in mind that the correlation structure in this case shows close agreement between men and women, I shall first present the result of the analysis for the total group. An examination of the eigenvalues indicates that three factors are sufficient to explain the common variance. Table 22 gives both the unrotated and the rotated factor matrices with estimated communalities.

A study of the rotated factor matrix shows that the first factor has its highest loadings in the variables 3:6, 3:7 and 3:2. Thus, it can apparently be regarded as measuring working capacity primarily from the qualitative point of view. It is responsible for almost half of the total contribution of variance. The second factor includes the variables 3:3, 3:4 and 3:5, so that it may be interpreted as measuring adjustment to work. Finally, the third factor has its highest loadings in

3:8, 3:1 and 3:2 and in all likelihood measures the capacity for work primarily from the quantitative point of view.

If we pay special attention to the prognosis-group and percentage-group criteria, we find, as expected, that the former variable has its highest loading in the quality factor, and the latter in the quantity factor. However, the factor structure in general shows that both the variables are complex measures of working capacity and that the differences between them cannot be said to be more than a difference of degree.

I have also carried out corresponding factor analyses among men and women separately and arrived thereby at results which can be interpreted in the same way as those now reported in the total group. In the female group, however, the $T_K$ values indicate that I here have reason to assume a fourth factor, which has high loadings only in the prognosis-group and percentage-group criteria. This confirms what I previously noted, namely, that the women do not display the same high correlations as the men between these two general rating criteria and the objective performance measurements in the testing programme.

Thus, the result of the completed factor analysis may be said to confirm the picture of the structure of the criterion variables which I obtained by a detailed examination of the intercorrelations. I was able to show that by a combination of measurements of performance and systematic observations it is possible to discern three different aspects of working capacity in mentally retarded persons, namely, a quantitative, a qualitative and an aspect concerning motivation and adjustment to work situations. In this context, it should be remembered that in an earlier section I tried to approach the problem of the dimensionality of working capacities by means of factor analysis of the quantitative performances in eight tasks. The question now arises, whether the results of the two analyses may be said to point in the same direction. A comparison of the outcome of the analyses and a study of the correlations between the variables in question here show that this is the case.

In the first completed factor analysis I found reason to distinguish between one factor with high loadings in the assembling of electrical panels and the making of wooden frames, and a second factor with high loadings in the cutting of threads in nuts, assembling of brackets and embroidering of cat motifs. This result agrees to some extent with the possibility of isolating a qualitative aspect of behaviour at work, as justified by the second factor analysis. We already know from the section on measurement of qualitatively determined performances that the assembly of electrical panels and the making of wooden frames may be described as rather difficult and demanding tasks, while the opposite applies to the cutting of threads in nuts and the assembling of brackets. As I have already had reason to remark, this pattern does not fit as closely when we consider the embroidering of cat motifs, which, according to our measurements of qualitative performance, is one

*Table 23.* Product-moment correlations between certain composite criterion measures and quantitative performances in selected tasks. Total group. ($n$=64).

| Variable | Quality | | | Adjustment | | |
|---|---|---|---|---|---|---|
| | 3:2 | 3:6 | 3:7 | 3:3 | 3:4 | 3:5 |
| Assembling electrical panels | .73 | .74 | .79 | .25 | .50 | .23 |
| Making wooden frames | .65 | .65 | .75 | .33 | .51 | .34 |
| Threading nuts | .38 | .52 | .56 | .61 | .68 | .35 |
| Assembling brackets | .28 | .33 | .42 | .48 | .52 | .25 |
| Embroidering cat motifs | .48 | .46 | .51 | .19 | .50 | .23 |
| Pressing eyelets | .52 | .47 | .64 | .24 | .47 | .09 |
| Turning holders | .67 | .55 | .68 | .29 | .50 | .13 |
| Packing blocks | .60 | .62 | .72 | .27 | .50 | .18 |

of the most difficult and demanding of the tasks. The type of difficulty encountered by the workers in the sewing task seems, however, to be rather specific and different from the corresponding difficulties in the other tasks. This is suggested by the correlation survey given in Table 23, which shows more or less unambiguously that the three criterion variables, which I found to measure the qualitative factor of working capacity, consistently show lower correlations with the cutting of threads in nuts, the assembling of brackets *and* the embroidering of cat motifs, in comparison with the assembling of electrical panels and the making of wooden frames. This means that the two last-mentioned tasks, to a greater extent that the other three, involve the factor of qualitative working capacity.

Moreover further examination of correlations shows that it is also possible to prove a certain—if by no means as obvious—relationship between the criterion variables which define the third factor of working capacity—covering motivation to work and adjustment in work situations—and the classification of the tasks which emerged as a result of the first factor analysis. As shown in Table 23, the differences here go in the opposite direction, with the result that the three criterion variables show a certain tendency to a higher correlation with the cutting of threads in nuts and the assembling of brackets than with the assembling of electrical panels and the making of wooden frames. The tendency becomes the more obvious, however, only as regards the one of the criterion variables, and, in this respect too, the embroidery of cat motifs falls outside the framework.

The results I have now reported are thus in certain respects more tentative than definitive. However, they suggest that the two approaches I used to analyse the problem of the structure of the capacities for work may support and complement each other. They are also signs that "performance is determined jointly by cognition (know-how, habit, or what has been learned) and motivation" (Siegel 1968). In the relatively harder and more complicated tasks the work result is governed primarily by the qualitative proficiency, while the degree of motivation

is of greater significance in the relatively easier, short-cycle or assembly-line tasks (cf. Holm and Palmér 1972).

Now that I have concluded the analysis of the structure of the criterion variables, I am ready, in the rest of the investigation, to study the possibility of predicting these different aspects of working capacity among the mentally retarded.

# Chapter V

# Prediction of Performance at Work and Adjustment to Work

In the foregoing pages I have given, *inter alia,* an account of the selection of prediction variables. The basic hypothesis, which I there justified in detail, involved the assumption that working capacity and adjustment to work in the mentally retarded can be predicted from their performance in a number of ability variables, mainly in the sphere of psychomotor function but also in those of spatial-inductive and numerical function, and from observations of behaviour in the environment offered by sheltered workshops.

In the next section I showed how I carried out measurements, with a satisfactory degree of reliability, of 27 different ability variables, intended for use as predictors, and how I could use systematic observations of behaviour to construct a total of nine different variables, which measure personality characteristics in wide functional areas and which I also intended to use as predictors. Finally, in the immediately preceding section I similarly analysed the criterion problem in detail and described the derivation of a number of criterion variables which measure different aspects of work performance and adjustment to work in the mentally retarded.

Before I now go on to the actual prediction studies, I shall describe the temporal sequence of the collection of data, which I have previously mentioned only *en passant.* Here we can distinguish two stages, the first of which covers mainly the years 1967–8. During this period I first performed the measurements of ability described above. Once these were complete, I collected the values for the three criterion variables which I designated prognosis-group, percentage-group and wages-group criteria and which were based on the supervisors' observations in connection with the regular work on the shop floor. Thus the necessary temporal conditions existed in which I was able to analyse the possibilities of predicting rated working capacity from measurements of psychomotor, spatial-inductive and numerical abilities. I shall present such an analysis in the next section.

In the second stage of the investigations, covering the years 1969–71, I began by carrying out the systematic observations of behaviour, which resulted in the construction of nine prediction variables. I then started the series of

measurements of performances in eight tasks, which gave me objective criteria of working capacity. In this connection, I carried out detailed observations of the subjects' behaviour in the standardized work situations; I was then able to analyse these observations and make them into criteria of adjustment to work.

The temporal sequence thus reported, shows that the pre-requisite conditions exist for analysing the possibilities of predicting all the criteria which I obtained in connection with the measurements of performance in the eight tasks, partly from the measurements of ability and partly from the nine prediction variables, which measure characteristic features of working behaviour. Such an analysis follows in a later section of this chapter.

## 1. Prediction of Rated Working Capacity

**A. Product-moment correlations between ability variables and rating criteria**

I shall begin by studying Table 24, which gives the calculated product-moment correlations between the prediction and criterion variables discussed here. As the table shows, I have here also included the wages-group criterion, notwithstanding that I previously announced my intention to omit this variable from further discussion. I merely wish to prove that in general the wages-group and percentage-group criteria are similar, as regards their correlations with the prediction variables. This fact is a further indication that in terms of content they are similar and that I shall not therefore lose much information by omitting the wages-group criterion in the future.

If we thus concentrate our attention on the prognosis-group and percentage-group criteria, we find that, in the case of the men, both display significant correlations with all the prediction variables. The magnitude of the coefficients varies between .33 and .70, and shows on the whole close agreement between the two criteria, which will also be confirmed by a study of the diagram in Fig. 20. Here it must be remembered that certain of the prediction variables are evaluated in such a way that we may expect negative values. This applies to nos. 5—6, 8—9, 11, 18 and 24—27. I have disregarded this fact in the diagram and have removed the minus signs.

So far as the women are concerned, in some cases they too display significant and marked correlations between the prediction variables and the two criteria. However, the magnitude of these correlations is more or less consistently lower than among the men. This is particularly true as regards the percentage-group criterion, which is also shown in Fig. 20. In fact the situation seems to be that all the numerical, as well as the majority of the spatial-inductive and even certain psychomotor prediction variables, reveal negligible correlations with this criterion among the women.

The extent to which the women display lower correlations than the men

*Table 24.* The correlations of the ability variables with the rating criteria.

| Prediction variable | Prognosis group | | Percentage group | | Wages group | |
|---|---|---|---|---|---|---|
| | Men $n$=47 | Women $n$=33 | Men $n$=47 | Women $n$=33 | Men $n$=39 | Women $n$=24 |
| 1. Crawford screws | .42** | .35* | .53** | .40* | .32* | .33 |
| 2. Crawford pins | .56** | .58** | .60** | .55** | .42** | .49* |
| 3. O'Connor test | .64** | .45** | .70** | .41* | .40* | .31 |
| 4. Mark-making | .61** | .38* | .51** | – .01 | .45** | .02 |
| 5. Wire labyrinth | – .58** | – .23 | – .60** | – .22 | – .49** | .04 |
| 6. Threading | – .54** | – .31 | – .43** | – .36* | – .44** | – .27 |
| 7. Plates | .53** | .56** | .57** | .37* | .33* | .55** |
| 8. MTTA | – .64** | – .52** | – .66** | – .49** | – .43** | – .40* |
| 9. Tapping | – .49** | – .55** | – .49** | – .32 | – .40* | – .53** |
| 10. Copying | .56** | .52** | .60** | .48** | .36* | .41* |
| 11. Reaction time | – .50** | – .47** | – .41** | – .22 | – .50** | – .32 |
| 12. Memory for design | .33* | .23 | .35* | .22 | .32* | .34 |
| 13. Surface development | .46** | .44** | .42** | .42* | .20 | .27 |
| 14. Block design | .37** | .35* | .45** | .22 | .19 | .29 |
| 15. Raven's matrices | .52** | .27 | .47** | .27 | .29 | .18 |
| 16. The Leiter Scale | .61** | .47** | .57** | .24 | .35* | .47* |
| 17. Figure classification | .54** | .27 | .60** | – .04 | .39* | .04 |
| 18. Concepts | – .57** | – .34* | – .54** | – .14 | – .32* | – .22 |
| 19. Addition, UB | .45** | .17 | .42** | .04 | .50** | .29 |
| 20. Addition, W | .67** | .37* | .54** | .08 | .56** | .26 |
| 21. Subtraction, W | .57** | .29 | .47** | .10 | .51** | .23 |
| 22. Highest number | .58** | .36* | .57** | .22 | .46** | .21 |
| 23. Identical forms | .50** | .26 | .52** | .13 | .35* | .12 |
| 24. Vocabulary | – .38** | – .46** | – .36* | – .24 | .04 | – .33 |
| 25. Reaction time, S.D. | – .47** | – .42* | – .45** | – .22 | – .36* | – .34 |
| 26. Wire labyrinth, S.D. | – .54** | – .18 | – .58** | – .02 | – .43** | .15 |
| 27. Tapping, S.D. | – .36* | – .39* | – .39** | – .22 | – .25 | – .41* |

** indicates a value significant at the 1% level.
* indicates a value significant at the 5% level.

between the prediction variables and the two criteria is more directly illustrated by the diagram in Fig. 21. This confirms that the difference is comparatively marked. I showed above that there is a corresponding sexual difference, as regards the magnitude of the intercorrelations between the prediction variables. Since I am not convinced that the circumstances mentioned are, in their entirety, caused by the lower variability of the female group, it seems advisable to treat men and women as two separate groups in the following regression analysis.

To sum up, from the correlation coefficients in Table 24, I am able to establish that my hypothesis as to the possibilities of predicting criteria of working capacity on the basis of ability variables receives strong support as long as the male experimental group is considered. It would seem that psychomotor, spatial-

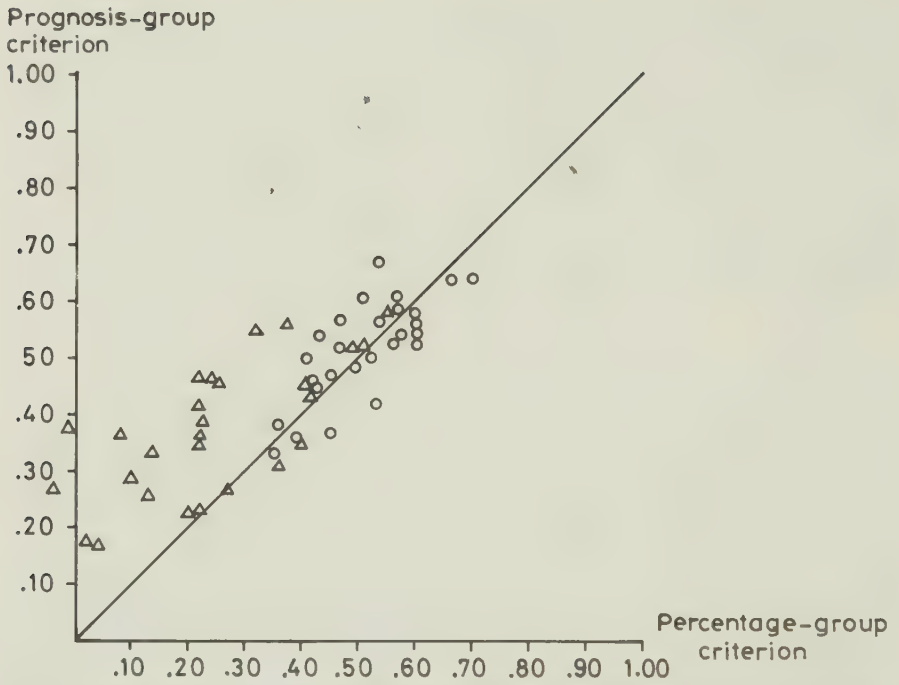



Fig. 20. Correlations of the ability variables with the prognosis-group and percentage-group criteria respectively among men (○) and women(△).

inductive and numerical variables can serve as more or less equivalent predictors. The vocabulary test also shows a significant correlation with the criteria, though on a comparatively low level. Finally, regarding the variability measurements, the coefficients in Table 24 suggest that they too may be used as prediction variables. As I have already said, the situation is somewhat more ambiguous as regards the female experimental group.

### B. Prediction of rating criteria from combinations of ability variables

*How to find "the best regression equation"*

However, I am not only, or even primarily, interested in the correlations between separate test variables and criteria. There is the more important problem of how I can so combine different prediction variables that together they lead to the best possible prediction of working capacity in the mentally retarded. I first investigated the immediate possibility of improving the prediction by adding up the scores on certain obvious prediction variables and correlating the total with the criteria. The results of these efforts were not particularly encouraging, however, and I therefore decided to use other lines of action.

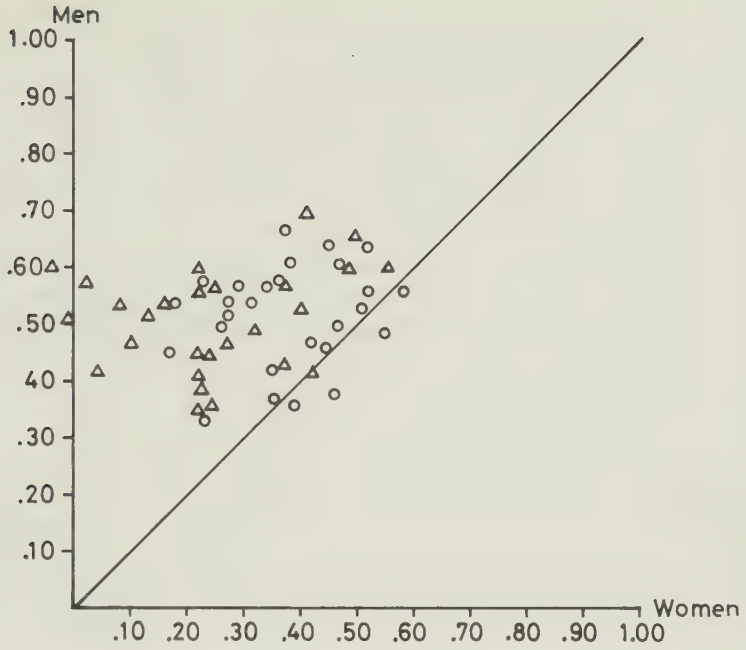



Fig. 21. Correlations of the ability variables with the prognosis-group criterion (○) and the percentage-group criterion (△) among men and women.

In the next step the possibility arose of combining prediction variables and calculating their total prediction capacity by multiple-regression analysis. This posed the problem of limiting the number of predictors and of selecting "the best regression equation". As various authors maintain (Draper and Smith 1966, Lord and Nowick 1968) there is no obviously correct solution to that problem. Certain scholars have recommended, and to some extent also empirically tested, an approach which presupposes the existence of a basic factor structure in both predictors and criteria (Henrysson 1957, Guilford 1965, Burket 1964). This approach to the problem will probably be fruitful in the long run, but can only come into its own in large scale studies. Therefore, I have in my investigation found convincing reasons in favour of the application of stepwise multiple-regression analysis, using separate test scores as independent variables. Even if it is true that this procedure, despite its ingenuity, does not give a full guarantee that the best array of predictors will be obtained, it may be regarded from several points of view as the most satisfactory of those available.

*Pitfalls in stepwise multiple-regression analysis*

The results of stepwise multiple-regression analysis may, however, be deceptive in more than one respect, as handbooks on the subject demonstrate in detail (Guilford 1965, McNemar 1969, Nunnally 1967). In calculating the multiple correlation coefficient by applying the principle of least squares, one takes advantage of sampling errors among correlations, and thus obtains a higher value in a small sample than one would have obtained in a larger one. This bias, which is usually described as "shrinkage", may be of considerable extent in investigations with few individuals and a number of independent variables. Nevertheless it is possible to correct for it, and thus obtain an unbiased estimate of the population multiple correlation by applying the formula

$$\hat{R}^2 = 1 - (1 - R^2)\,(N - 1/N - k)$$

The problems of significance in multiple regression also call for special attention. As McNemar (1969) explains in detail, traditional methods of testing the significance of a multiple correlation coefficient include unsatisfactory approximations. For this reason, there has been a general change-over to the application for this purpose of an analysis of variance technique, which here involves the determination of the ratio between the mean square due to regression and the mean square due to residual terms, and which treats this ratio as an $F$ variate. Thus, I avoid the approximations mentioned above, but, in view of the possibilities of using chance to my own advantage in several respects, I must here choose a low risk level to ensure that the regression equation will function as an useful prediction instrument.

Besides the significance test just mentioned, which refers to the overall regression equation, it is also necessary to test the individual regression coefficients for significance, using the same technique of analysis of variance, In this connection, one often arrives at apparently contradictory results, the most common being that the multiple correlation coefficient is found to be significant perhaps by a wide margin, while one or more of the individual regression coefficients does not reach a significant level.

Only during recent years, apparently, have statisticians begun to analyse the relationships between these different $F$ tests and to seriously investigate the problem of how to deal with mutually contradictory results of the kind given above (Geary and Leser 1968, Cramer 1972). In these studies, I have taken the view that it should be possible to use a computer programme, which proceeds in a stepwise fashion, but that under no circumstances should a regression equation so obtained be accepted unless it fulfils a twofold condition. The overall $F$ test should show that the multiple correlation coefficient reaches a statistically significant level by an ample margin, and the partial $F$ tests should yield values which

are significant at $\alpha=0.05$. How large a margin should be demanded as regards the overall $F$ test can only be determined from case to case.

Even if, as suggested above, corrections are made for "shrinkage" and the results of completed significance analyses are studied with care, yet another circumstance—and that perhaps the most important—must be taken into consideration. The usual procedure in stepwise regression analysis involves the selection from a large number of variables of a smaller number of variables which have in common the quality of showing high correlations with a criterion. Yet occasionally these correlations have only become high as a result of sampling errors, and will therefore be lower in new samples. In estimating the multiple correlation coefficient, I capitalize in this sense on chance. This bias may be of considerable magnitude and I cannot correct for it by applying a statistical formula. Instead I should test the selected variables with their calculated regression coefficients in a new sample. This procedure is described as cross-validation and must be regarded as an imperative necessity in all prediction studies using multiple-regression analysis to reach more definitive results. I have already indicated that in this investigation I intend to carry out cross-validations so far as resources allow.

*The results of the analyses*

Having described my attitude to the complications and limitations of the method, I am ready to report the results of the series of stepwise multiple-regression analyses which I performed in the context discussed here. In order that the number of independent variables should not be greater than necessary, I decided from the start to eliminate "Memory for design" and "Subtraction, W". "Memory for design" as a test has two defects. It produces an obviously skewed distribution in the experimental group and its reliability is not quite satisfactory. Moreover the test shows lower correlations with the criteria than other spatial-inductive tests. "Subtraction, W" was omitted because its correlation with "Addition, W" is as high as .85. The remaining 25 prediction variables served as independent variables in the first analyses, which were performed with two computer programmes, Stepwise Multiple Regression (Bäckström 1970) and BMD 02R Stepwise Regression.

As Guilford (1965, p. 415) and other authors point out, it is seldom worth including more than a few variables in a final test battery for the prediction of a certain criterion. As a rule, further addition of variables does not appreciably improve the prediction. Instead, there is a danger that the element of error variance in the predicted values will increase. As will be apparent from the above account, I allowed the number of variables entered into the regression equations to be strictly determined by the significance limits fixed above. This meant that the

*Table 25.* Results of stepwise multiple-regression analyses of the prognosis-group criterion.

| Experimental group | Predictor variables | Coefficients | Std. Errors | *F* to remove | Zero order correlations | Indices of prediction accuracy | |
|---|---|---|---|---|---|---|---|
| Men, *n*=47 | 20. Addition, W | 0.02415 | 0.00740 | 10.6596 | .67 | Multiple R | .90 |
| | 3. O'Connor test | 0.02530 | 0.00676 | 13.9880 | .64 | Adjusted for Df | .88 |
| | 11. Reaction time | −0.00770 | 0.00217 | 12.6171 | −.50 | *F*-value | 23.41 |
| | 14. Block design | −0.13401 | 0.02435 | 30.2770 | .37 | Df | 7/39 |
| | 15. Raven's matrices | 0.07648 | 0.02257 | 11.4787 | .52 | Std Error of est. | .52 |
| | 6. Threading | −0.00226 | 0.00067 | 11.4266 | −.54 | Adjusted for Df | .56 |
| | 13. Surface development | 0.04297 | 0.01872 | 5.2722 | .46 | Prop. explained variance | .81 |
| Women, *n*=33 | 2. Crawford pins | 0.07838 | 0.02635 | 8.8482 | .58 | Multiple R | .68 |
| | 9. Tapping | −0.00361 | 0.00138 | 6.8084 | −.55 | Adjusted for Df | .67 |
| | | | | | | *F*-value | 12.82 |
| | | | | | | Df | 2/30 |
| | | | | | | Std Error of est. | .75 |
| | | | | | | Adjusted for Df | .77 |
| | | | | | | Prop. explained variance | .46 |

*Table 26.* Results of stepwise multiple-regression analyses of the percentage-group criterion.

| Experimental group | Predictor variables | Coefficients | Std Errors | *F* to remove | Zero order correlations | Indices of prediction accuracy | |
|---|---|---|---|---|---|---|---|
| Men, *n*=47 | 3. O'Connor test | 0.04407 | 0.00930 | 22.4795 | .70 | Multiple R | .84 |
| | 17. Figure classification | 0.09585 | 0.02775 | 11.9276 | .60 | Adjusted for Df | .82 |
| | | | | | | *F*-value (Df 5/41) | 9.47 |
| | 14. Block design | −0.11528 | 0.03384 | 11.6014 | .45 | Std Error of est. | .74 |
| | 10. Copying | 0.08935 | 0.03517 | 6.4536 | .60 | Adjusted for Df | .78 |
| | 25. Reaction time, S.D. | −0.01818 | 0.00867 | 4.3932 | .45 | Prop. explained variance | .70 |
| Women, *n*= 33 | 2. Crawford pins | 0.9986 | 0.03385 | 8.7038 | .55 | Multiple R | .71 |
| | 17. Figure classification | −0.09689 | 0.03141 | 9.5130 | −.04 | Adjusted for Df | .69 |
| | 10. Copying | 0.14653 | 0.05526 | 7.0310 | .48 | *F*-value (Df 3/29) | 9.81 |
| | | | | | | Std Error of est. | .85 |
| | | | | | | Adjusted for Df | .88 |
| | | | | | | Prop. explained variance | .50 |

numbers varied between one and seven. Tables 25 and 26 report the results of stepwise multiple-regression analyses of the prognosis-group and the percentage-group criteria respectively among men and women.

The values in the tables seem to provide convincing confirmation of my hypothesis as to the possibilities of successfully predicting working capacity in the mentally retarded from their performances in a number of ability variables. This conclusion applies irrespective of which criterion we are considering.

Once the multiple correlation coefficients have been corrected for the number of degrees of freedom, they amount for the men to .88 (the prognosis-group criterion) and .82 (the percentage-group criterion) respectively. In the female group the corresponding figures are .67 and .69. The statistical significance in all these coefficients is verified by the given $F$ values, which in the male group exceed the 1 % level by 7.4 and 5.6 times respectively. In the female group these values exceed the 1 % level by 2.4 and 2.2 times respectively. The difference between men and women which we found for correlations between separate test variables and criteria thus appears in a characteristic fashion also in these multiple-regression analyses. In the female group we can note not only lower multiple correlation coefficients but also lower $F$ values.

If we consider the tables as regards the question of which test variables have entered into the regression equations, we find that both psychomotor and spatial-inductive and numerical variables are represented. Some further observations are worthy of special note. So far as the women are concerned, almost only psychomotor variables are included, which is not surprising if we remember that in their case correlations between other variables and criteria are comparatively weak or negligible. However, it is interesting to note that "Figure Classification" is included as a suppressor variable in the analysis referring to the percentage-group criterion. The "Block Design" test fulfils the same function in the regression analyses regarding both the criteria in the male group.

Of the numerical variables only one, "Addition W," is represented at all. This is the case in the regression analysis of the prognosis-group criterion among men. This fact can also be related to the correlations between separate test variables and criteria. Thus, Table 24 shows that the numerical test variables are almost invariably more highly correlated with the prognosis-group criterion than with the percentage-group criterion. At the same time, we find that the reverse is true with regard to several of the psychomotor variables, and to "Block Design" and "Figure Classification". The differences noted would seem to be due to the fact that to some extent the prognosis-group and percentage-group criteria cover different aspects of working capacity.

*The relative importance of the different predictor groups*

It would be of interest to estimate more accurately the relative importance, as

regards the prediction, of the individual test variables or groups of variables. Guilford (1965, p. 399) has described a method of performing such an estimation, which involved the use of the β-coefficients to calculate how much the different predictors contribute to the reduction of the variance in the criterion by the formula

$$R^2_{1.23} = \beta_{12.3} r_{12} + \beta_{13.2} r_{13.}$$

However, as Guilford also points out, such a method of calculation presupposes that all β-*r* products are positive. In other words, it is impossible to apply this procedure if suppressor variables have entered into the regression equations, which we have just found to be the case in the analyses reported above.

In order to obtain further information by another means concerning the predicitve power offered by combinations of similar tests, I chose to carry out separate multiple-regression analyses. Thus, I first included tests nos. 1—11, all of which measure psychomotor abilities *sensu lato,* separately as independent variables. Their places were then taken by the remaining tests, which—as we already know—are spatial-inductive, numerical, or—in one case—verbal in nature. In this connection the measures of intra-subject variability were left out of consideration. I still held fast to the principle of accepting as predictors only the variables or groups of variables which had entered into the equations with statistically significant regression coefficients (at the 5 % level). The results are reported in summary form in Tables 27 and 28.

A study of the tables justifies the conclusion that tests in the area of psychomotor functions in general are the best variables to use for predicting the working capacity of the mentally retarded. This is suggested by both the multiple correlation coefficients and their *F* values. Support for this conclusion emerges most clearly in the female experimental group, which displays higher multiple correlation coefficients when the prediction is made from psychomotor-test variables, as regards both the prognosis-group and percentage-group criteria. In the male experimental group there is a congruent difference only in the case of the percentage-group criterion. The *F* values are considerably higher in all cases in which the psychomotor tests are used as prediction variables, which confirms that they contain less error variance than other variables.

Figure 22 gives a visual illustration of certain of the conditions just described. Here I elected to express the accuracy of the predictions as proportions of explained variance in the criteria. A striking feature is the low degree of success in predicting working capacity in women from spatial-inductive and numerical ability tests. Moreover, it is quite obvious that in general it is possible appreciably to increase the accuracy of the prediction by combining psychomotor with spatial-inductive and numerical prediction variables.

*Table 27.* Prediction of rating criteria from psychomotor abilities.

| Criteria | Experimental groups | Predictor variables | Multiple correlation | Prop. explained variance | Degrees of freedom | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|---|
| Prognosis-group criterion | Men, $n$=47 | 3. O'Connor test<br>4. Mark-making<br>11. Reaction time | .76 | .57 | 3/43 | 19.27 | 4.5 |
| " | Women, $n$=33 | 2. Crawford pins<br>9. Tapping | .68 | .46 | 2/30 | 12.82 | 2.4 |
| Percentage-group criterion | Men, $n$=47 | 3. O'Connor test<br>8. MTTA | .74 | .56 | 2/44 | 27.79 | 5.4 |
| " | Women, $n$=33 | 2. Crawford pins | .55 | .30 | 1/31 | 13.42 | 1.8 |

*Table 28.* Prediction of rating criteria from spatial-inductive and numerical abilities.

| Criteria | Experimental groups | Predictor variables | Multiple correlation | Prop. explained variance | No. of degrees of freedom | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|---|
| Prognosis-group criterion | Men, $n$=47 | 20. Addition, W<br>16. The Leiter Scale<br>14. Block design<br>13. Surface development | .76 | .57 | 4/42 | 13.97 | 3.7 |
| " | Women, $n$=33 | 16. The Leiter Scale | .47 | .22 | 1/31 | 8.60 | 1.1 |
| Percentage-group criterion | Men, $n$=47 | 17. Figure classification<br>18. Concepts | .64 | .41 | 2/44 | 15.43 | 3.0 |
| " | Women, $n$=33 | 13. Surface development | .42 | .17 | 1/31 | 6.47 | – |

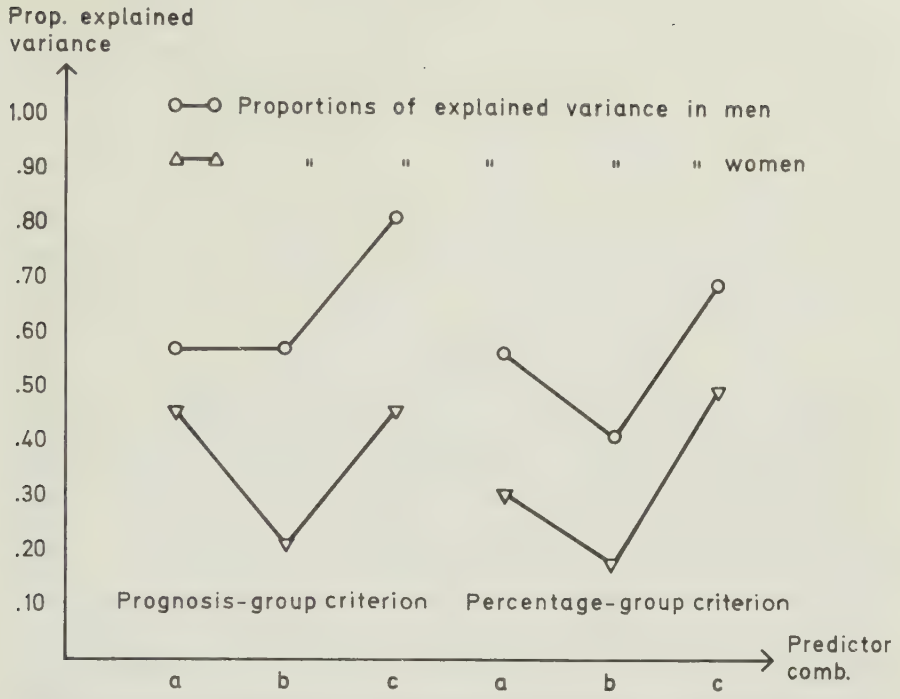



Fig. 22. Proportions of explained criterion variance in the prediction of rating criteria from 3 different combinations of predictors:
a) Combination of psychomotor abilities; b) Combination of spatial-inductive and numerical abilities; c) Combination of a + b.

### C. Cross-validation studies

The authors of the critical surveys reviewed earlier of prognosis studies among the mentally retarded referred repeatedly to the absence of cross-validation investigations. The use of multiple-regression analyses increases the need for such investigations, since these analyses involve a definite risk of taking advantage of chance, a risk which was discussed in the preceding section. Thus, I had good reason to use different possibilities of cross-validating my own results.

Such an opportunity for cross-validation arose when the Malmöhus County Council decided to carry out an industrial psychological investigation of the pupils who were about to conclude their compulsory education at the special school. As one member of my research project was involved in this investigation, I was able to influence the selection of test variables to roughly coincide with my own practice in my main investigation. I was also able to obtain the type of rating criterion which I described above as the prognosis-group criterion.

For detailed information on the composition of the experimental group in Malmöhus County, the reader is referred to a special account by Bodin and Willquist (1972). This indicates that the group—which consisted of 40 men and

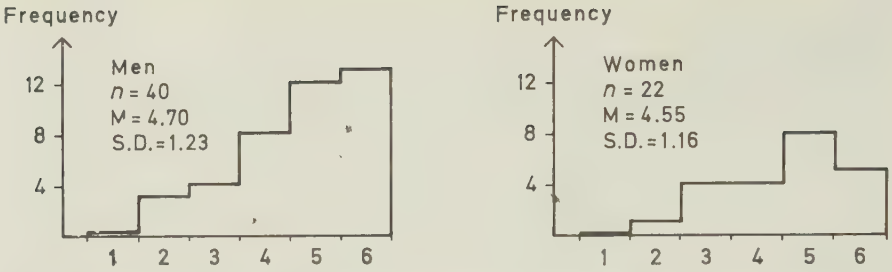



Fig. 23. Distributions of the prognosis-group criterion among men and women in the cross-validation group.

22 women—corresponds fairly well as regards general intellectual attainment and previous schooling with my own experimental group. On the other hand, the average age was as low as 18.4 years, which is to be compared with 25.3 years in the original group. The lower age means that members of the cross-validation group had distinctly less training in practical tasks. Naturally it also followed that the raters of criteria had had fewer opportunities to observe their dexterity and general working capacity. Nor had these raters as long experience of working conditions and standards as our supervisors at the sheltered workshop in Uppsala. Although the conditions for using the group in Malmöhus County for cross-validation purposes were thus not entirely favourable, I considered that they were acceptable and should be used.

The majority of pupils were assessed by three different teachers and, as in the main investigation, I allowed the average of their ratings to stand as the final score in the prognosis-group criterion. The report just mentioned showed that the reliability of these values was satisfactory. Their distribution is shown in the diagrams in Fig. 23. A scrutiny of these diagrams reveals that the distributions are negatively skewed, and that their average value is higher than the equivalent in my own experimental group. I interpreted this as an expression of excessive optimism on the part of the raters.

The actual cross-validation was performed by calculating the predicted criterion value for each individual in the Malmö group by inserting individual test results into the regression equation and by multiplying them by the regression coefficients, which I had calculated in my own experimental group and which are given in Table 25 above. On correlating these predicted criterion values with the observed values in the prognosis-group criterion I obtained the cross-validation coefficients given in Table 29.

As I had reason to expect, these cross-validation coefficients proved to be appreciably lower than the multiple correlation coefficients reported above. Nevertheless, they are all significant at the 1 % level and—regarded as validity coefficients—are comparatively high. One may here recall, for example, that

*Table 29.* Cross-validation coefficients referring to the prognosis-group criterion in the Malmö-group.

| Men, $n$=40 | Women, $n$=22 | Men and women, $n$=62 |
|---|---|---|
| .50 | .57 | .48 |

Bemis (1968) carried out a survey of the median validity of batteries of GATB aptitudes and thereby found that calculated φ-coefficients displayed median values which varied from .40 to .45.

However, I felt a need to cross-validate the results by applying yet another method. The main reason was my hesitation concerning the prognosis-group rating in the Malmö investigation. I showed above that this rating resulted in a negatively skewed criterion variable, with a comparatively high mean value. An examination of its correlations with the prediction variables indicates that these display several divergencies, compared with the corresponding data in the Uppsala investigation. These circumstances gave me reason to assume that the assessors in the Malmö investigation were working from a somewhat different frame of reference, which may have resulted in a steeper fall of the cross-validation coefficients in relation to the original multiple correlation coefficients than would otherwise have been the case. Another reason for trying to cross-validate the results by applying a new procedure arose from my failure to obtain any percentage-group rating whatsoever in the Malmö sample.

The new procedure which I therefore decided to use, has been discussed by, *inter alios,* Guilford (1954) and Horst (1966). The former calls it "double cross-validation", while the latter uses the designation "sample fractionation". As these terms suggest, the procedure involves the division of the experimental group into two parts, and the derivation of a set of weights in each. In my case, I divided the men in the Uppsala sample into two groups. To one group were assigned 24 individuals with odd identity numbers in the school and training groups respectively, while the remaining 23 made up the other group. Stepwise multiple-regression analyses were then performed of both the prognosis-group and the percentage-group criteria in both sub-groups, with the results given in Tables 30 and 31.

A study of the tables shows that the calculated multiple correlation coefficients vary between .71 and .91. The $F$ values exceed the 1 % level about five times as regards the prognosis-group criterion, and some three times in the case of the percentage-group criterion. The composition of the prediction variables giving these high values alternates between sub-groups and criteria, which is a natural consequence of the fact that a number of independent variables correlate almost as highly with the criteria.

Using the same procedure as described above for the Malmö group, I calculated predicted criterion values in the two sub-groups and correlated them

*Table 30.* Results of stepwise multiple-regression analyses of the prognosis-group criterion in two male sub-samples.

| Experimental group | Predictor variables | Coefficients | Std Errors | *F* to remove | Zero order correlations | Indices of prediction accuracy | |
|---|---|---|---|---|---|---|---|
| Sub-sample 1 | 8. MTTA | −0.00179 | 0.00029 | 35.5273 | −.83 | Multiple R | .91 |
| *n*=24 | 9. Tapping | −0.00333 | 0.00156 | 4.5308 | −.56 | Adjusted for Df | .90 |
| | 14. Block design | −0.10121 | 0.03244 | 9.7367 | .25 | *F*-value (Df 4/19) | 22.80 |
| | 13. Surface development | 0.06662 | 0.02881 | 5.3469 | .50 | Std Error of est. | .51 |
| | | | | | | Adjusted for Df | .55 |
| | | | | | | Prop. explained variance | .83 |
| Sub-sample 2 | 20. Addition, W | 0.03746 | 0.01200 | 9.7528 | .81 | Multiple R | .89 |
| *n*=23 | 3. O'Connor test | 0.04805 | 0.01495 | 10.3248 | .80 | Adjusted for Df | .88 |
| | 14. Block design | −0.06959 | 0.02827 | 6.0580 | .50 | *F*-value (Df 3/20) | 23.73 |
| | | | | | | Std Error of est. | .54 |
| | | | | | | Adjusted for Df | .56 |
| | | | | | | Prop. explained variance | .79 |

*Table 31.* Results of stepwise multiple-regression analyses of the percentage-group criterion in two male sub-samples.

| Experimental group | Predictor variables | Coefficients | Std Errors | *F* to remove | Zero order correlations | Indices of prediction accuracy | |
|---|---|---|---|---|---|---|---|
| Sub-sample 1 | 8. MTTA | −0.00193 | 0.00041 | 22.6298 | −.71 | Multiple R | .71 |
| *n*=24 | | | | | | *F*-value (Df 1/22) | 22.63 |
| | | | | | | Std Error of est. | .93 |
| | | | | | | Adjusted for Df | .97 |
| | | | | | | Prop. explained variance | .51 |
| Sub-sample 2 | 3. O'Connor test | 0.03925 | 0.01878 | 4.3684 | ..78 | Multiple R | .87 |
| *n*=23 | 25. Reaction time, S.D. | −0.05031 | 0.01768 | 8.0926 | −.55 | Adjusted for Df | .84 |
| | 15. Raven's matrices | −0.08221 | 0.03717 | 4.8916 | .53 | *F*-value (Df 4/18) | 13.64 |
| | 10. Copying | 0.12868 | 0.04979 | 6.6781 | .68 | Std Error of est. | .68 |
| | | | | | | Adjusted for Df | .73 |
| | | | | | | Prop. explained variance | .75 |

*Table 32.* Cross-validation coefficients referring to the prognosis-group and percentage-group criteria in the two sub-samples.

| Prognosis-group criterion | | Percentage-group criterion | |
|---|---|---|---|
| Sub-sample 1 $n$=24 | Sub-sample 2 $n$=23 | Sub-sample 1 $n$=24 | Sub-sample 2 $n$=23 |
| .57 | .68 | .62 | .74 |

with those observed. Thus, I obtained the cross-validation coefficients given in Table 32.

An examination of this table shows that the cross-validation coefficients calculated as described above seem to be of greater magnitude than the corresponding data in the Malmö sample. The maximum figure in this sample, namely, .57, is here a minimum, while other coefficients are on a considerably higher level. In some measure, the reason may simply be that the groups are smaller, but otherwise it seems probable that the difference results from the fact that the criterion ratings in this case were made by the same assessors and from the same frame of reference.

In conclusion, I may, in any case, assert that, taken together, the cross-validation coefficients just reported provide good evidence of the possibilities of successfully predicting rated working capacity in the mentally retarded on the basis of ability variables.

## 2. Prediction of Performance and Adjustment in Eight Tasks

Having analysed in the previous section the possibilities of predicting the prognosis-group and percentage-group criteria, I shall now pass on to illustrate the corresponding problems which arise when the criteria consist of objectively determined performances at work and observations of the worker's behaviour in the working environment. I shall first examine the possibilities of predicting the previously described quantitative performances in eight selected tasks from measurements of ability and observations of behaviour. I shall then make a corresponding analysis with respect to the composite criterion measures.

### A. Prediction of quantitative performances in eight tasks

*Measurements of ability as predictors.* Here let us first study Tables 33 and 34,

*Table 33.* The correlations of the ability variables with quantitative performances in eight tasks. Men ($n$=40).

| Prediction variable | Pressing eyelets | Threading nuts | Turning holders | Assembly of brackets | Assembly of electrical panels | Making wooden frames | Embroidering cat motifs | Packing blocks |
|---|---|---|---|---|---|---|---|---|
| 1. Crawford screws | .43** | .27 | .30 | .33* | .70** | .64** | .42** | .26 |
| 2. Crawford pins | .56** | .31* | .39* | .12 | .50** | .49** | .48** | .29 |
| 3. O'Connor test | .61** | .41** | .48** | .39* | .61** | .60** | .62** | .46** |
| 4. Mark-making | .38* | .49** | .50** | .16 | .54** | .46** | .33* | .51** |
| 5. Wire labyrinth | –.45** | –.49** | –.51** | –.35* | –.54** | –.55** | –.43** | –.46** |
| 6. Threading | –.42** | –.25 | –.32* | –.28 | –.41** | –.36* | –.28 | –.21 |
| 7. Plates | .50** | .48** | .47** | .33* | .62** | .57** | .53** | .62** |
| 8. MTTA | –.49** | –.49** | –.49** | –.19 | –.56** | –.49** | –.45** | –.50** |
| 9. Tapping | –.44** | –.54** | –.44** | –.44** | –.45** | –.45** | –.40** | –.49** |
| 10. Copying | .51** | .31* | .49** | –.02 | .32* | .35* | .39* | .43** |
| 11. Reaction time | –.36* | –.41** | –.39* | –.25 | –.50** | –.45** | –.37* | –.37* |
| 12. Memory for design | .40** | .31* | .19 | .07 | .18 | .09 | .37* | .26 |
| 13. Surface development | .36* | .14 | .20 | –.09 | .22 | .22 | .29 | .20 |
| 14. Block design | .45** | .27 | .22 | .03 | .33* | .36* | .34* | .42** |
| 15. Raven's Progressive Matrices (coloured version) | .49** | .45** | .31* | .17 | .47** | .48** | .46** | .54** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 16. The Leiter International Scale | .46** | .42** | .43** | .21 | .55** | .56** | .52** | .62** |
| 17. Figure classification | .51** | .46** | .43** | .26 | .52** | .48** | .46** | .64** |
| 18. Concepts | −.34* | −.35* | −.44** | −.14 | −.42** | −.41** | −.33* | −.46** |
| 19. Addition, UB | .37* | .51** | .43** | .25 | .22 | .29 | .36* | .57** |
| 20. Addition, W | .41** | .52** | .46** | .19 | .29 | .35* | .51** | .57** |
| 21. Subtraction, W | .34* | .47** | .49** | .20 | .16 | .28 | .43** | .58** |
| 22. Highest number | .41** | .51** | .41** | .22 | .54** | .59** | .40** | .66** |
| 23. Identical forms | .34* | .37* | .36* | .05 | .30 | .34* | .35* | .54** |
| 24. Vocabulary | −.26 | −.25 | −.29 | −.07 | −.21 | −.21 | −.14 | −.33* |
| 25. Reaction time, S.D. | −.34* | −.46** | −.43** | −.34* | −.49** | −.49** | −.34* | −.41** |
| 26. Wire labyrinth, S.D. | −.47** | −.48** | −.60** | −.33* | −.57** | −.57** | −.46** | −.52** |
| 27. Tapping, S.D. | −.38* | −.43** | −.37* | −.26 | −.37* | −.29 | −.28 | −.30 |

** indicates a value significant at the 1% level and * a value significant at the 5% level.

*Table 34.* The correlations of the ability variables with quantitative performances in eight tasks. Women ($n$=24).

| Prediction variable | Pressing eyelets | Threading nuts | Turning holders | Assembly of brackets | Assembly of electrical panels | Making wooden frames | Embroidering cat motifs | Packing blocks |
|---|---|---|---|---|---|---|---|---|
| 1. Crawford screws | .65** | .24 | .44* | .45* | .77** | .66** | .33 | .47* |
| 2. Crawford pins | .33 | –.02 | .47* | .05 | .43* | .47* | .34 | .41* |
| 3. O'Connor test | .45* | .06 | .32 | .18 | .53** | .53** | .21 | .42* |
| 4. Mark-making | .37 | –.02 | .35 | –.07 | .46* | .40* | .02 | .23 |
| 5. Wire labyrinth | –.21 | –.12 | –.27 | .11 | –.56** | –.47* | –.12 | –.20 |
| 6. Threading | –.44* | –.26 | –.09 | –.26 | –.40* | –.46* | –.18 | –.23 |
| 7. Plates | .58** | .10 | .57** | .32 | .75** | .73** | .36 | .37 |
| 8. MTTA | –.72** | –.22 | –.52** | –.34 | –.70** | –.62** | –.45* | –.67** |
| 9. Tapping | –.38 | .15 | –.49* | –.17 | –.54** | –.54** | –.20 | –.57** |
| 10. Copying | .67** | .24 | .61** | .24 | .71** | .69** | .27 | .63** |
| 11. Reaction time | –.38 | –.08 | –.36 | –.22 | –.61** | –.67** | –.08 | –.20 |
| 12. Memory for design | .06 | .19 | .37 | –.09 | .41* | .41* | .11 | .44* |
| 13. Surface development | .47* | .07 | .28 | .19 | .37 | .39 | .14 | .38 |
| 14. Block design | .14 | .31 | .44* | .05 | .46* | .24 | .06 | .32 |
| 15. Raven's Progressive Matrices (coloured version) | .24 | .29 | .59** | .19 | .45* | .40* | .18 | .33 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 16. The Leiter International Scale | .27 | .13 | .33 | .00 | .36 | .32 | .13 | .44* |
| 17. Figure classification | .34 | −.05 | .46* | −.12 | .52** | .56** | −.04 | .28 |
| 18. Concepts | −.37 | .07 | −.42* | −.08 | −.37 | −.24 | −.12 | −.52** |
| 19. Addition, UB | .10 | .15 | .52** | .23 | .35 | .33 | .28 | .47* |
| 20. Addition, W | .14 | .04 | .42* | .02 | .37 | .33 | −.05 | .48* |
| 21. Subtraction, W | .11 | .17 | .37 | .02 | .25 | .31 | .00 | .38 |
| 22. Highest number | .14 | .15 | .35 | .04 | .25 | .24 | −.01 | .37 |
| 23. Identical forms | .20 | −.34 | .43* | −.03 | .24 | .09 | −.18 | .20 |
| 24. Vocabulary | −.41* | −.11 | −.36 | −.13 | −.39 | −.34 | .11 | −.55** |
| 25. Reaction time, S.D. | −.29 | .09 | −.37 | −.22 | −.40* | −.32 | −.09 | −.22 |
| 26. Wire labyrinth, S.D. | −.10 | .11 | −.37 | .20 | −.59** | −.43* | .03 | −.07 |
| 27. Tapping, S.D. | −.19 | .27 | −.60** | −.07 | −.51* | −.41* | −.28 | −.32 |

** indicates a value significant at the 1% level and * a value significant at the 5% level.

which give the product-moment correlations between the scores achieved in the ability variables and the quantitative performances in the eight tasks among men and women respectively. As I pointed out earlier (cf. p. 100), we have to bear in mind that some of the variables have been adjusted, so that positive correlations are expressed by negative values.

A glance at Table 33 reveals that in the male group substantial correlations appear to a large extent between prediction variables and criteria. A dozen of these correlations reach orders of magnitude between .60 and .70. Thus far, the facts seem to show a fairly good agreement with what we found in the analysis of the corresponding data referring to ability variables and rating criteria. Since the quantitative performances are more specific criteria, however, it is not surprising that here we get a more differentiated picture when we examine the correlation coefficients more closely. It turns out that 75 % of these coefficients are significant at either the 1 % or the 5 % level. From this point of view, they are distributed as shown in Table 35. A striking feature is that, in the case of the men, seven of the eight tasks show a similar state of affairs—a large number of significant and a small number of non-significant coefficients. However, the assembly of brackets deviates from this pattern. In most cases, the correlations between this task and the prediction variables are weak and non-significant.

If we pass on to the female group and consider Table 34, we find—as we found in studying the same question with regard to the rating criteria—that the number of significant correlations is much smaller than in the male group (only 32 % of the total number). From the point of view of significance, they are also distributed in a very characteristic way, as may be ascertained by studying the columns in Table 35 referring to the female group. Here we find that the great majority of the significant correlations are located in the assembly of electrical

*Table 35.* Distribution of zero-order correlation coefficients from the point of view of significance.

| Criterion variables | No. of coefficients which are significant at the 1% level | | significant at the 5% level | | not significant | |
|---|---|---|---|---|---|---|
| | Men | Women | Men | Women | Men | Women |
| Pressing eyelets | 17 | 4 | 9 | 4 | 1 | 19 |
| Threading nuts | 17 | 0 | 5 | 0 | 5 | 27 |
| Turning holders | 16 | 6 | 6 | 8 | 5 | 13 |
| Assembling brackets | 1 | 0 | 6 | 1 | 20 | 26 |
| Assembling electrical panels | 17 | 10 | 3 | 8 | 7 | 9 |
| Making wooden frames | 16 | 8 | 5 | 8 | 6 | 11 |
| Embroidering cat motifs | 14 | 0 | 9 | 1 | 4 | 26 |
| Packing blocks | 19 | 5 | 2 | 7 | 6 | 15 |
| Total | 117 | 33 | 42 | 37 | 54 | 146 |

panels, the making of wooden frames, the turning of holders and the packing of blocks—precisely the four tasks which were found to be referrable to one and the same working-capacity factor in the previously reported factor analysis. On the contrary, the correlations between the ability variables and the cutting of threads in nuts, the assembly of brackets and the embroidery of cat motifs—the tasks which in the factor analysis displayed the highest loadings in the second working-capacity factor—are almost exclusively non-significant. The remaining task—the pressing of eyelets—occupies an intermediate position from this point of view, just as it did in the factor analysis. Thus, it appears as if the prediction possibilities in the female group are to a very large extent dependent on the kind of working capacity which it is desired to predict.

Having examined the zero-order correlations in this way and found that a considerable number of them are of such an order of magnitude that it seems to be possible to reach a satisfactory accuracy of prediction both in the male group and—though to a more limited extent—in the female group, I now want to explore these possibilities in greater detail by considering some stepwise multiple-regression analyses. The results of these analyses, with reference to the quantitative performances in all eight tasks, will be found in Tables 36 and 37.

If we first consider the results in the male group, we may note in Table 36 that the multiple correlation coefficients vary between .44 and .88. Not surprisingly, the assembly of brackets has the lowest value, which is, of course, a consequence of the fact that most of the correlations between this task and the prediction variables were found to be non-significant. Again, the assembly of electrical panels (.88) and the making of wooden frames (.80) show the highest values. Here we may recall that it was these two tasks which had the highest loadings in the first working-capacity factor among both men and women. In the other five tasks, the multiple correlation coefficients are bunched together, with a range of variation from .65 to .73 and with *F*-values which exceed the 1 % level from 2.6 to 4.0 times. If we disregard the assembly of brackets, we may accordingly state, by way of summary, that in the male group there are satisfactory possibilities of predicting the quantitative performances in the tasks from measurements of ability.

As might be expected after the above examination of the zero-order correlations, the prediction possibilities in the female group prove to vary noticeably from one task to another, a feature which may be studied in Table 37. No regression equation can be accepted as satisfactory either in cutting threads in nuts or in embroidering cat motifs. In the former task, no predictor variable reaches a statistically significant level and in the latter, the overall *F*-value falls below the 1 % level. It is true that, in the assembly of brackets, the stepwise multiple-regression analysis has resulted in a comparatively high multiple correlation coefficient, but the significance of this may be questioned, as the *F*-value is low and

*Table 36.* Prediction of quantitative performances in eight tasks from measurements of ability. Men ($n$=40).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Pressing eyelets | 3 O'Connor test<br>17 Figure classification | .68 | .46 | 2/37 | 15.67 | 3.0 |
| Threading nuts | 9 Tapping<br>25 Reaction time, S.D. | .65 | .42 | 2/37 | 13.66 | 2.6 |
| Turning holders | 26 Wire labyrinth, S.D.<br>20 Addition, W | .66 | .43 | 2/37 | 14.16 | 2.7 |
| Assembly of brackets | 9 Tapping | .44 | .19 | 1/38 | 9.08 | 1.2 |
| Assembly of electrical panels | 1 Crawford screws<br>4 Mark-making<br>17 Figure classification<br>14 Block design<br>25 Reaction time | .88 | .77 | 5/34 | 23.26 | 6.4 |
| Making wooden frames | 1 Crawford screws<br>22 Highest number<br>26 Wire labyrinth, S.D. | .80 | .64 | 3/36 | 21.53 | 4.9 |
| Embroidering cat motifs | 3 O'Connor test<br>20 Addition, W | .68 | .46 | 2/37 | 15.65 | 3.0 |
| Packing blocks | 22 Highest number<br>26 Wire labyrinth, S.D. | .73 | .53 | 2/37 | 20.99 | 4.0 |

exceeds the 1 % level by only 1.4 times. Consequently, the standard error in the predicted values will be high and the usefulness of the regression equation will be limited. The quantitative performances in the assembly of electrical panels and the making of wooden frames, on the other hand, may be predicted with evident success, as the figures in Table 37 show. The regression equations obtained in the remaining tasks—pressing eyelets, turning holders and packing blocks—may also be regarded as satisfactory.

Thus, in the female group we have found clear indications that the prediction possibilities vary according to the nature of the criterion. If this criterion consists of quantitative performances, with high loadings in the first working-capacity factor, the prediction possibilities seem to be good. On the other hand, if we wish to predict performances representing the second working-capacity factor, we shall have less success. Even among the men, there is, in fact, a tendency for the prediction possibilities to vary in a similar way, but this tendency is not as marked in that group.

If, finally, we inquire which ability variables contribute most to the predictions

*Table 37.* Prediction of quantitative performances in eight tasks from measurements of ability. Women ($n$=24).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Pressing eyelets | 8 MTTA | .72 | .51 | 1/22 | 23.31 | 2.9 |
| Threading nuts | – a) | – | – | – | – | – |
| Turning holders | 10 Copying<br>19 Addition, UB | .71 | .51 | 2/21 | 10.89 | 1.9 |
| Assembly of brackets | 1 Crawford screws<br>5 Wire labyrinth<br>6 Threading<br>4 Mark-making<br>19 Addition, UB | .79 | .62 | 5/18 | 5.89 | 1.4 |
| Assembly of electrical panels | 1 Crawford screws<br>9 Tapping<br>14 Block design<br>2 Crawford pins | .92 | .84 | 4/19 | 25.27 | 5.6 |
| Making wooden frames | 7 Plates<br>10 Copying | .82 | .68 | 2/21 | 22.06 | 3.8 |
| Embroidering cat motifs | 8 MTTA | .45 | .20 | 1/22 | 5.52 | – |
| Packing blocks | 8 MTTA<br>19 Addition, UB<br>24 Vocabulary | .82 | .67 | 3/20 | 13.42 | 2.7 |

a) No predictor variable reaches a statistically significant level.

reported, we may note that psychomotor tests play a leading part. However, among both men and women, spatial-inductive and numerical variables also occur. In the female group the vocabulary test also contributes to the prediction of the performance in packing blocks. In conclusion, it is worth noting that one or another of the variability measurements enters into four of the eight regression equations in the male group. On the whole, however, there are grounds for stating that there is good agreement with the predictor structure which emerged in the prediction of the prognosis-group and percentage-group criteria.

*Observational data as predictors.* I shall now go on to study the possibilities of predicting quantitative performances in the eight tasks from the nine prediction variables which I developed on the basis of observations of behaviour. Here I may recall that six of these variables were measured by the same number of sub-scales in a general observation form of the check-list type, while the remaining three derive from a five-degree rating schedule (cf. p. 47 above and Appendix B:1-2). The calculated product-moment correlations between these prediction variables

*Table 38.* The correlations of the observation variables with quantitative performances in eight tasks. Men ($n$=40).

| Predictor variables | Pressing eyelets | Threading nuts | Turning holders | Assembly of brackets | Assembly of electrical panels | Making wooden frames | Embroidering cat motifs | Packing blocks |
|---|---|---|---|---|---|---|---|---|
| 1 Emotional stability | −.18 | −.26 | −.20 | −.27 | −.38* | −.37* | −.30 | −.24 |
| 2 Ability to co-operate | .10 | −.04 | .05 | −.14 | −.03 | −.08 | −.05 | .05 |
| 3 Ability to communicate through speech | −.17 | −.04 | −.19 | −.07 | −.25 | −.18 | −.04 | −.08 |
| 4 Motivation to work | −.26 | −.49** | −.35* | −.32* | −.35* | −.44** | −.42** | −.50** |
| 5 Movement co-ordination and working pace | −.41** | −.44** | −.46** | −.27 | −.53** | −.46** | −.33** | −.50** |
| 6 Technical knowledge and skills | −.57** | −.63** | −.57** | −.36* | −.62** | −.62** | −.53** | −.57** |
| 7 Rated work proficiency | .63** | .54** | .61** | .31* | .69** | .65** | .55** | .64** |
| 8 Rated adjustment to work | .50** | .57** | .39* | .37* | .48** | .51** | .67** | .48** |
| 9 Rated quantitative performance | .57** | .58** | .55** | .59** | .57** | .54** | .46** | .54** |

Note. Predictor variables 1–6 in Tables 38 and 39 have been adjusted so as to yield negative correlations with positive characteristics. **indicates a value significant at the 1% level and * a value significant at the 5% level.

and the quantitative performances among men and women respectively will be found in Tables 38 and 39.

If we examine the values in the male group, we may note that the variables with emotional contents, such as "emotional stability" and "ability to co-operate" and also "ability to communicate through speech", display, as a rule, insignificant correlations with the criteria. On the other hand, the picture is a different one, when we consider the variables "technical knowledge and skills" and "movement co-ordination and working pace". Both these variables display, with one or two exceptions, substantial correlations with the criteria, which are generally also significant at the 1 % level. The same thing applies to variables 7–9, which are ratings of work proficiency, adjustment to work and quantity of work—all made with reference to the five-degree rating schedule. Finally, the

*Table 39.* The correlations of the observation variables with quantitative performances in eight tasks. Women ($n$=24). (See note in Table 38).

| Predictor variables | Pressing eyelets | Threading nuts | Turning holders | Assembly of brackets | Assembly of electrical panels | Making wooden frames | Embroidering motifs | Packing cat blocks |
|---|---|---|---|---|---|---|---|---|
| 1 Emotional stability | −.08 | −.22 | .17 | −.12 | −.07 | .02 | −.06 | .07 |
| 2 Ability to co-operate | .10 | −.37 | .21 | −.30 | .14 | .18 | −.14 | .27 |
| 3 Ability to communicate through speech | −.25 | −.19 | −.38 | −.11 | −.39 | −.29 | −.40* | −.23 |
| 4 Motivation to work | .06 | −.36 | .27 | −.37 | .24 | .09 | −.17 | −.03 |
| 5 Movement co-ordination and working pace | −.44* | −.36 | −.25 | −.29 | −.46* | −.43* | −.34 | −.58** |
| 6 Technical knowledge and skills | −.17 | −.34 | −.17 | −.12 | −.20 | −.25 | −.08 | −.37 |
| 7 Rated work proficiency | .39 | .26 | .48* | .15 | .61** | .61** | .35 | .66** |
| 8 Rated adjustment to work | .26 | .52** | .00 | .56** | .07 | .02 | .33 | .14 |
| 9 Rated quantitative performance | .42* | .36 | .13 | .54** | .03 | .08 | .31 | .30 |

"motivation to work" variable is also correlated to a considerable extent with the criteria. Thus, there seem to be comparatively good prospects of predicting the quantitative performances in the male group from these nine observation variables.

In the female group, significant correlations between the observation variables and the criteria are much fewer and farther between, as Table 39 shows. The absence of such correlations is characteristic not only of variables 1 and 2, which have emotional contents, but also of "motivation to work" and "technical knowledge and skills". The significant correlations which occur are chiefly to be found in the "movement co-ordination and working pace" and "rated work proficiency" variables. Thus, judging solely from the zero-order correlations, the prediction possibilities appear to be considerably smaller in the female group than in the male.

That this really is the case is confirmed when we look at the results of the stepwise multiple-regression analyses, which will be found in Tables 40 and 41.

However, these two tables also contain other information of great interest. In the male group, the multiple correlation coefficients vary between .59 and .81, and the calculated *F*-values reach in all cases a satisfactory order of magnitude, in that they exceed the 1 % level by between 2.8 and 3.6 times. The prediction variables entered into the equations with statistically significant regression coefficients derive both from the general observation form of check-list type and from the five-degree rating schedule. The conclusion from these analyses is that the quantitative working performances of mentally retarded men can be predicted not only by ability variables—as I showed earlier—but also by observation variables. This is intrinsically not very surprising. However, I had scarcely ex-

*Table 40.* Prediction of quantitative performances in eight tasks from observations of behaviour. Men ($n$=40).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Pressing eyelets | 7 Rated work proficiency | .63 | .40 | 1/38 | 25.50 | 3.5 |
| Threading nuts | 6 Technical knowledge and skills | | | | | |
| | 9 Rated quantitative performance | .68 | .47 | 2/37 | 16.21 | 3.1 |
| Turning holders | 7 Rated work proficiency | .61 | .37 | 1/38 | 22.66 | 3.1 |
| Assembly of brackets | 9 Rated quantitative performance | .59 | .35 | 1/38 | 20.47 | 2.8 |
| Assembly of electrical panels | 7 Rated work proficiency | | | | | |
| | 1 Emotional stability | .74 | .55 | 2/37 | 22.91 | 4.4 |
| Making wooden frames | 7 Rated work proficiency | | | | | |
| | 1 Emotional stability | .71 | .50 | 2/37 | 18.82 | 3.6 |
| Embroidering cat motifs | 8 Rated adjustment to work | | | | | |
| | 7 Rated work proficiency | | | | | |
| | 5 Movement co-ordination and working pace | | | | | |
| | 1 Emotional stability | | | | | |
| | 2 Ability to co-operate | .81 | .65 | 5/34 | 12.70 | 3.5 |
| Packing blocks | 7 Rated work proficiency | | | | | |
| | 4 Motivation to work | .71 | .51 | 2/37 | 18.91 | 3.6 |

pected that the observation variables would hold their own so well as predictors alongside the ability variables. If we make a comparative study of Tables 36 and 40 and in the process observe both the magnitudes of the multiple correlation coefficients and the *F*-values, we shall find, in fact, that the observation variables predict the "cutting of threads in nuts", the "assembly of brackets" and the "embroidering of cat motifs" better than the ability variables, while the opposite is the case with the "assembly of electrical panels", the "making of wooden frames" and the "packing of blocks". The fact that the tasks are distributed in this particular way is, of course, not a coincidence but is a reflection of the fact that here we have a fresh example—this time in the male group—of a relationship between the results of the previous factor analysis of the quantitative performances at work and the predictability of these performances. From the prediction point of view, this observation is of extraordinary interest, since it means that observation variables and ability variables should complement each other and together make

*Table 41.* Prediction of quantitative performances in eight tasks from observations of behaviour. Women ($n$=24).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Pressing eyelets | 5 Movement co-ordination and working pace | .44 | .19 | 1/22 | 5.22 | – |
| Threading nuts | 8 Rated adjustment to work | .52 | .27 | 1/22 | 8.16 | 1 |
| Turning holders | 7 Rated work proficiency | .48 | .23 | 1/22 | 6.52 | – |
| Assembly of brackets | 8 Rated adjustment to work<br>1 Emotional stability<br>6 Technical knowledge and skills<br>7 Rated work proficiency | .77 | .59 | 4/19 | 6.97 | 1.5 |
| Assembly of electrical panels | 7 Rated work proficiency | .61 | .37 | 1/22 | 13.17 | 1.7 |
| Making wooden frames | 7 Rated work proficiency | .61 | .38 | 1/22 | 13.36 | 1.7 |
| Embroidering cat motifs | – a) | | | | | |
| Packing blocks | 7 Rated work proficiency | .66 | .43 | 1/22 | 16.74 | 2.1 |

a) No predictor variable reaches a statistically significant level.

it possible to predict performances successfully within wide and heterogeneous fields of work.

If we then pass on to consider the results of the stepwise multiple-regression analyses in the female group, we may note in Table 41 that, as regards "pressing eyelets", "turning holders" and "embroidering cat motifs", the regression equations are quite unsatisfactory, since the multiple correlation coefficients are remarkably low and the *F*-values in all cases are below the 1 % level. The situation is better, as regards the remaining five tasks. Here the multiple correlation coefficients vary between .52 and .77, and the corresponding *F*-values exceed the 1 % level by between 1 and 2.1 times. This means that in these cases the prediction possibilities may be described as perhaps acceptable but not particularly good. If we compare the present results of the analysis with the corresponding results using ability variables as predictors, we may note a certain improvement only in the prediction of quantitative performances in "cutting threads in nuts". In all other cases, the predictions from ability variables are either equivalent, as in the "assembly of brackets" and the "embroidering of cat motifs", or better, as in "pressing eyelets", "turning holders", "assembling electrical panels", "making wooden frames" and "packing blocks". Thus, it is noteworthy that the observation variables are not equally good predictors in the female group as in the male and that therefore they cannot supplement the ability variables to the same extent as we found with regard to the male group. Here we may recall that we have already noted a parallel state of affairs to this in connection with our examination of the correlations between the prognosis-group and the percentage-group criteria, on the one hand, and the quantitative performances in the eight tasks, on the other. This gives further support to the assumption which I expressed in this connection—that subjective judgements of women contain more complicating elements than subjective judgements of men.

*Combinations of measurements of ability and observational data as predictors.* After the account given in the previous subsection of the predictability of quantitative performances from measurements of ability and observational data separately, it remains to investigate to what extent these predictions can be improved by combining the most promising variables from the two predictor groups. For this purpose, I have made fresh stepwise multiple-regression analyses, in which the independent variables were just such combinations. The results are shown in Tables 42 and 43. In addition, I have drawn the diagrams reproduced in Figs. 24 and 25, in order to facilitate a comparison between the predictive power of different predictor combinations.

If we first consider the results in the male group, we find that the previously expressed expectation—that the observation and ability variables might supplement each other and together yield improved predictions—is confirmed in seven of the eight tasks. The exception is the "assembly of electrical panels", in which even

*Table 42.* Prediction of quantitative performances in eight tasks from a combination of measurements of ability and observations of behaviour. Men ($n$=40).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Pressing eyelets | Rated work proficiency<br>O'Connor test | .72 | .51 | 2/37 | 19.60 | 3.7 |
| Threading nuts | Technical knowledge and skills<br>Tapping | .69 | .48 | 2/37 | 16.95 | 3.2 |
| Turning holders | Rated work proficiency<br>Wire labyrinth, S.D. | .66 | .44 | 2/37 | 14.40 | 2.8 |
| Assembly of brackets | Rated quantitative performance<br>Tapping<br>Rated work proficiency<br>Reaction time | .75 | .57 | 4/35 | 11.56 | 3.0 |
| Assembly of electrical panels | Crawford screws<br>Rated work proficiency<br>Highest number<br>Block design | .86 | .74 | 4/35 | 24.97 | 6.4 |
| Making wooden frames | Rated work proficiency<br>Crawford screws<br>Highest number<br>Motivation to work<br>Movement co-ordination and working pace | .86 | .73 | 5/34 | 18.58 | 5.1 |
| Embroidering cat motifs | Rated adjustment to work<br>O'Connor test<br>Addition, W<br>Technical knowledge and skills<br>Reaction time | .87 | .76 | 5/34 | 21.88 | 6.0 |
| Packing blocks | Highest number<br>Motivation to work<br>Rated work proficiency | .79 | .63 | 3/36 | 20.58 | 4.7 |

the ability variables alone resulted in a high multiple correlation coefficient and an equally high $F$-value, which are not exceeded here. The largest increases in prediction ability arise as regards "assembly of brackets", "embroidery of cat motifs", "packing blocks" and "making wooden frames", while the changes as regards the three remaining tasks are fairly unimportant. This means that comparatively large differences remain between the different tasks, as regards the possibilities of predicting the quantitative performances in these tasks. Thus,

*Table 43.* Prediction of quantitative performances in eight tasks from a combination of measurements of ability and observations of behaviour. Women ($n$ = 24).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Pressing eyelets | MTTA<br>Crawford screws<br>Rated quantitative performance | .82 | .67 | 3/20 | 13.29 | 2.7 |
| Threading nuts | Rated adjustment to work | .52 | .27 | 1/22 | 8.16 | 1.0 |
| Turning holders | Copying<br>Addition, UB | .71 | .51 | 2/21 | 10.89 | 1.9 |
| Assembly of brackets | Rated adjustment to work<br>Crawford screws<br>Emotional stability<br>Technical knowledge and skills<br>Rated quantitative performance | .84 | .71 | 5/18 | 8.85 | 2.1 |
| Assembly of electrical panels | Crawford screws<br>Tapping<br>Block design<br>Crawford pins | .92 | .84 | 4/19 | 25.27 | 5.6 |
| Making wooden frames | Plates<br>Copying | .82 | .68 | 2/21 | 22.06 | 3.8 |
| Embroidering cat motifs | MTTA<br>Ability to communicate through speech<br>Motivation to work<br>Vocabulary | .78 | .61 | 4/19 | 7.43 | 1.7 |
| Packing blocks | MTTA<br>Addition, UB<br>Vocabulary | .82 | .67 | 3/20 | 13.42 | 2.7 |

Table 42 shows that the multiple correlation coefficients vary between .66 and .88 and that the $F$-values exceed the 1 % level by between 2.8 and 6.4 times. Apart from these variations, which are obviously connected with the varying natures of the tasks, there is reason to emphasize, in conclusion, that all the regression equations fulfil high requirements from the point of view of significance and that consequently they confirm, in a convincing manner, the possibilities of successfully predicting quantitative performances from measurements of ability and observational data.

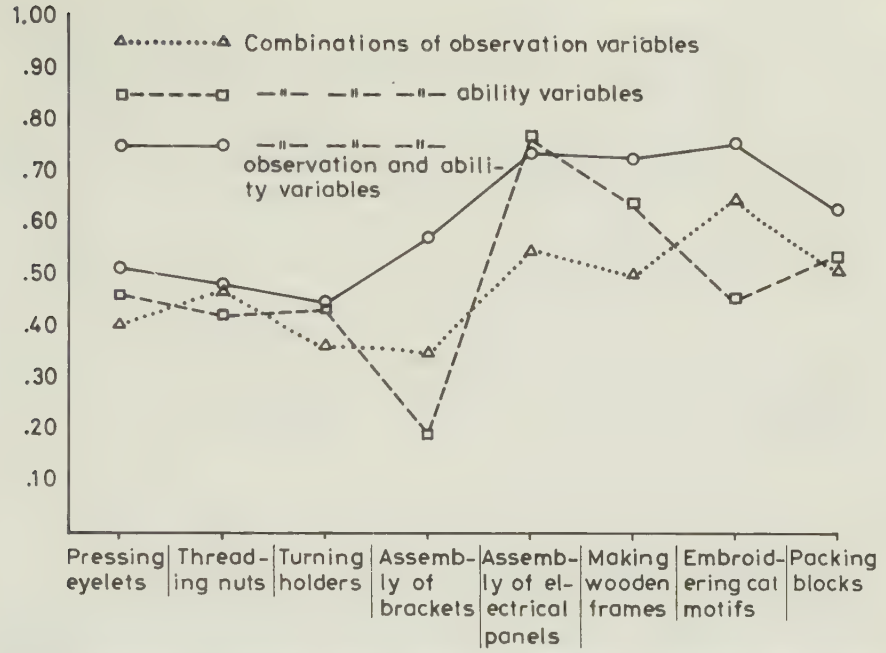



Fig. 24. Proportions of explained criterion variance in the prediction of quantitative performances in eight tasks from different combinations of predictors. Men ($n$=40).

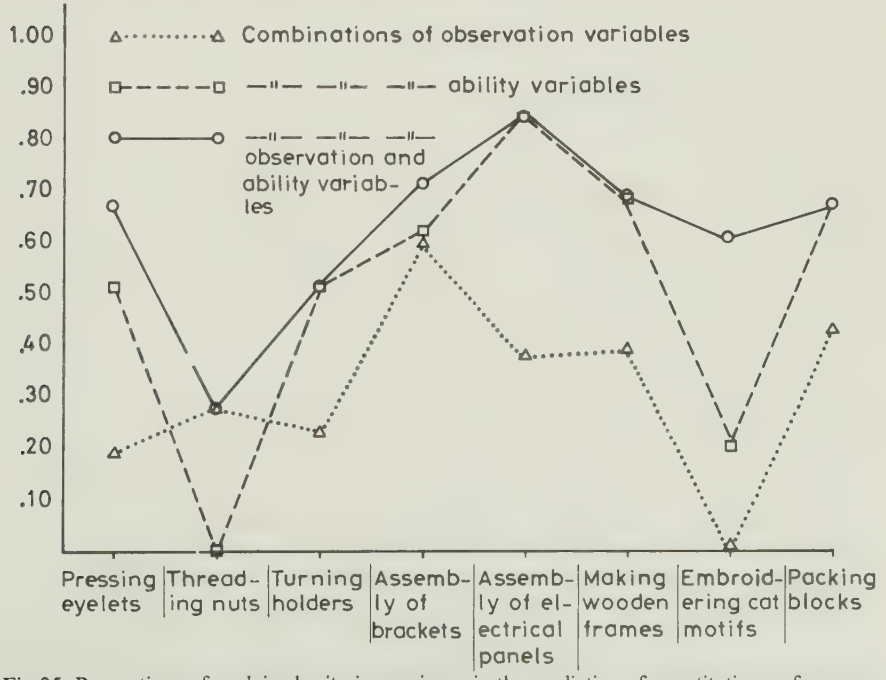



Fig 25. Proportions of explained criterion variance in the prediction of quantitative performances in eight tasks from different combinations of predictors. Women ($n$=24).

In the female group, the combinations of measurements of ability and observational data as predictors result only in less marked increases, as regards the proportions of explained criterion variance, as Fig. 25 shows. Such increases occur in three tasks, viz. "embroidering cat motifs", "pressing eyelets" and "assembling brackets". In the first case, the increase is very obvious, while in the two latter cases it is on a smaller scale. To certainly as great an extent as in the male group, there remain differences between the different tasks, as regards the possibilities of predicting performances in these tasks. If we disregard "cutting threads in nuts", we may state, by way of summary, that the prediction possibilities demonstrated also in the female group are fairly satisfactory. Several of the multiple correlation coefficients are high. However, one has to bear in mind that the group is a small one and that the degree of significance is smaller than in the male group.

**B. Prediction of composite criterion measures**

*Ability measurements as predictors.* In a previous section, I gave a detailed account of the genesis of a number of composite criterion measures based on measurements of performance and observations of behaviour in the eight tasks. I shall now show how I also studied the possibilities of predicting these composite criterion measures by examining zero-order correlations and carrying out stepwise multiple-regression analyses. In this connection also, I find it convenient to first consider the results of analyses in which the independent variables consisted exclusively of measurements of ability.

In Tables 44 and 45 will be found the correlations of the ability variables in men and women with the composite criterion measures, which are eight in number. As before, some prediction variables display negative correlations, in consequence of the method by which they were evaluated. The same applies to the criterion variables 3:2, 3:3, 3:5 and 3:6. I have previously stated that I regard the average quantitative performance as a main criterion, since it would seem to be primarily decisive of the individual's chances of asserting himself or herself in a workplace. The magnitudes of the correlations between this composite measure and the ability variables reflect, of course, the magnitudes of the correlations which we have previously found between the quantitative performances in the eight tasks and these same ability variables. Among the men, they vary between .27 and .64. Most of them are significant at the 1 % level, while the correlations with "memory for design", "surface development" and "vocabulary" are insignificant. Among the women, the correlations between the ability variables and the average quantitative performance are, as expected, much weaker in a large number of cases. Only 10 of them reach statistical significance at the 1 % or the 5 % level. However, some of them are comparatively high.

The correlations of the ability variables with eight composite criterion measures. Men ($n$=40).

| Prediction variable | 3:1 Quantitative performance | 3:2 Qualitative performance | 3:3 Adjustment to work, I | 3:4 Adjustment to work, II | 3:5 Total stoppage time | 3:6 Work proficiency, I | 3:7 Work proficiency, II | 3:8 Rated quantitative performance |
|---|---|---|---|---|---|---|---|---|
| 1. Crawford screws | .51** | −.52** | −.03 | .31* | −.30 | −.53** | .53** | .50** |
| 2. Crawford pins | .48** | −.67** | −.04 | .40** | −.25 | −.66** | .66** | .48** |
| 3. O'Connor test | .64** | −.60** | −.19 | .50** | −.25 | −.61** | .67** | .61** |
| 4. Mark-making | .51** | −.50** | .04 | .26 | −.08 | −.63** | .61** | .49** |
| 5. Wire labyrinth | −.58** | .66** | .12 | −.38* | .41** | .66** | −.62** | −.52** |
| 6. Threading | −.38* | .47** | −.17 | −.06 | .10 | .51** | −.46** | −.35* |
| 7. Plates | .63** | −.59** | .04 | .28 | −.02 | −.67** | .72** | .60** |
| 8. MTTA | −.55** | .75** | −.01 | −.32* | .11 | .82** | −.75** | −.53** |
| 9. Tapping | −.56** | .50** | .14 | −.31* | .36* | .55** | −.52** | −.50** |
| 10. Copying | .42** | −.60** | .05 | .30 | .04 | −.61** | .66** | .37* |
| 11. Reaction time | −.47** | .54** | .03 | −.33* | .02 | .62** | −.62** | −.58** |
| 12. Memory for design | .29 | −.25 | −.15 | .37* | −.05 | −.27 | .31* | .29 |
| 13. Surface development | .24 | −.39* | −.01 | .27 | .04 | −.37* | .44** | .27 |
| 14. Block design | .37* | −.44** | .04 | .20 | .10 | −.53** | .62** | .42** |
| 15. Raven's Progressive Matrices (coloured version) | .51** | −.42** | −.09 | .33* | .12 | −.52** | .66** | .53** |
| 16. The Leiter International Scale | .57** | −.64** | −.07 | .38* | −.04 | −.70** | .76** | .62** |
| 17. Figure classification | .57** | −.51** | −.02 | .31* | .05 | −.51** | .68** | .53** |
| 18. Concepts | .44** | −.52** | .01 | .26 | −.06 | −.64** | .65** | .47** |
| 19. Addition, UB | .46** | −.26 | −.18 | .23 | .01 | −.29 | .47** | .42** |
| 20. Addition, W | .50** | −.45** | −.18 | .31* | .06 | −.48** | .61** | .50** |
| 21. Subtraction, W | .45** | −.41** | −.29 | .37* | .08 | −.43** | .57** | .49** |
| 22. Highest number | .57** | −.47** | −.09 | .31* | −.11 | −.52** | .68** | .59** |
| 23. Identical forms | .40** | −.38* | .07 | .17 | .08 | −.51** | .60** | .38* |
| 24. Vocabulary | .27 | −.35* | .03 | .19 | −.04 | −.43** | .43** | .30 |
| 25. Reaction time, S.D. | −.50** | .41** | .18 | −.36* | .09 | .47** | −.56** | −.61** |
| 26. Wire labyrinth, S.D. | −.61** | .80** | .10 | −.41** | .29 | .77** | −.71** | −.57** |
| 27. Tapping, S.D. | −.41** | .45** | .06 | −.21 | .34* | .48** | −.42** | −.33* |

** indicates a value significant at the 1% level and * a value significant at the 5% level.

In the male group, the average qualitative-performance measurement shows quite as high correlations with the ability variables as the average quantitative-performance measurement. Only two of the coefficients are insignificant and the highest is as high as .75. Even in the female group, the quality measurement is more closely correlated with the ability variables than the quantity measurement

*Table 45.* The correlations of the ability variables with eight composite criterion measures. Women ($n$=24).

| Prediction variable | 3:1 Quantitative performance | 3:2 Qualitative performance | 3:3 Adjustment to work,I | 3:4 Adjustment to work,II | 3:5 Total stoppage time | 3:6 Work proficiency, I | 3:7 Work proficiency, II | 3:8 Rated quantitative performance |
|---|---|---|---|---|---|---|---|---|
| 1. Crawford screws | .67** | −.58** | −.21 | .15 | −.15 | −.58** | .50* | .50* |
| 2. Crawford pins | .42* | −.56** | .33 | −.12 | .19 | −.43* | .50* | .33 |
| 3. O'Connor test | .45* | −.37 | .00 | .10 | −.10 | −.38 | .52** | .43* |
| 4. Mark-making | .29 | −.28 | .31 | −.31 | .11 | −.17 | .23 | .20 |
| 5. Wire labyrinth | −.31 | .46* | −.07 | .02 | −.16 | .53** | −.53** | −.19 |
| 6. Threading | −.39 | .21 | .00 | .00 | .03 | .42* | −.38 | −.33 |
| 7. Plates | .64** | −.71** | −.02 | .11 | −.07 | −.66** | .66** | .58** |
| 8. MTTA | −.71** | .54** | −.04 | −.08 | .00 | .59** | −.70** | −.61** |
| 9. Tapping | −.46* | .64** | −.10 | −.04 | −.06 | .53** | −.65** | −.39 |
| 10. Copying | .68** | −.71** | −.07 | .14 | −.19 | −.49* | .57** | .46* |
| 11. Reaction time | −.44* | .60** | .03 | −.08 | −.03 | .57** | −.51* | −.35 |
| 12. Memory for design | .32 | −.51* | −.12 | .19 | −.26 | −.27 | .25 | .11 |
| 13. Surface development | .38 | −.31 | .05 | .09 | .08 | −.32 | .35 | .42* |
| 14. Block design | .34 | −.55** | −.32 | .45 | −.13 | −.47* | .44* | .21 |
| 15. Raven's Progressive Matrices (coloured version) | .45* | −.63** | −.30 | .48* | −.11 | −.45* | .45* | .34 |
| 16. The Leiter International Scale | .33 | −.53** | .00 | .20 | −.09 | −.49* | .48* | .26 |
| 17. Figure classification | .33 | −.74** | .16 | −.07 | .07 | −.42* | .44* | .15 |
| 18. Concepts | .35 | −.57** | .05 | .07 | .09 | −.39 | .51* | .22 |
| 19. Addition, UB | .41* | −.41* | −.15 | .29 | −.01 | −.44* | .38 | .48* |
| 20. Addition, W | .29 | −.60** | −.12 | .20 | .04 | −.45* | .36 | .27 |
| 21. Subtraction, W | .27 | −.43* | −.23 | .35 | −.09 | −.33 | .33 | .38 |
| 22. Highest number | .26 | −.45* | −.21 | .32 | −.24 | −.27 | .35 | .27 |
| 23. Identical forms | .10 | −.37 | .14 | .05 | .18 | −.10 | .21 | .09 |
| 24. Vocabulary | .37 | −.48* | −.14 | .26 | −.08 | −.32 | .43* | .28 |
| 25. Reaction time, S.D. | −.30 | .57** | −.18 | .01 | −.34 | .36 | −.40* | −.17 |
| 26. Wire labyrinth, S.D. | −.21 | .54** | −.14 | .03 | −.19 | .41* | −.44* | .02 |
| 27. Tapping, S.D. | −.36 | .65** | −.25 | .01 | −.23 | .36 | −.47* | −.21 |

** indicates a value significant at the 1% level and * a value significant at the 5% level.

and here the difference is much more conspicuous. As many as 22 out of 27 coefficients are significant, as against only 10 with regard to the quantity measurement. This difference between the quality and the quantity variable corresponds to the state of affairs which we found when we examined the correlations of the ability variables with the quantitative performances in the different tasks, viz. that these correlations were comparatively high, as regarded long-cycle and qualitatively exacting tasks, such as assembling electrical panels and making wooden frames, while, on the other hand, they were lower and, in the female group, insignificant practically throughout, as regarded short-cycle and simple tasks, such as cutting threads in nuts and assembling brackets. It may be that the performances in the latter kind of task are more dependent on the worker's degree of ambition and persistence than on his or her skill and that they are more difficult to predict precisely on that account, especially among the women.

It is possible to find a certain amount of support for the hypothesis that there may be a difference here between men and women, if we consider the correlations between the ability variables and the three composite criterion measures which refer to adjustment to work and total stoppage time. In a total of 81 coefficients, there is in the female group only a single one which reaches statistical significance, while as many as 19 do so in the male group. However, here also we must bear in mind what I have pointed out previously, viz. that such differences may, at least in part, be due to the smaller degree of variability in the female group.

If, finally, we consider the correlations between the ability variables and the three composite criterion measures which refer to "average rating of work proficiency in the tasks, I and II" and "average rating of quantitative performance in the tasks", we may observe that here obvious correlations appear which are throughout somewhat higher among the men than among the women.

The results of the stepwise multiple-regression analyses carried out in this connection will be found in Tables 46 and 47. If we first consider the male group, we find that the analysis of the average quantitative-performance measure results in a satisfactory regression equation. The multiple correlation coefficient amounts to .82 and the corresponding $F$-value exceeds the 1 % level by 4.7 times. Thus, the possibilities of predicting this criterion are comparatively good, in spite of the fact that it contains a certain degree of heterogeneity. The same verdict may be pronounced as regards criterion 3.8, "average rating of quantitative performance". Just as we might expect from the zero-order correlations, we may also note that the prediction possibilities just commented on are excelled by those with regard to criterion 3:2, "average qualitative performance", and criteria 3:6 and 3:7, both of which consist of ratings of the work proficiency demonstrated in carrying out the eight tasks. Thus, we may predict qualitative performances with greater success than quantitative performances. Finally, the figures in Table 46

confirm that ability variables are weak predictors, if one wishes to predict the motivation shown and the adjustment to the selected work situations. It is true that the regression equation as regards criterion 3:4, "average rating of adjustment to work, II", is fairly satisfactory but not to the same extent as in other criteria.

In the female group, the accuracy of the predictions is less practically throughout, but the regression equations obtained may nevertheless be regarded as comparatively satisfactory, if we disregard the adjustment criteria 3:3 and 3:5.

*Table 46.* Prediction of eight composite criterion measures from measurements of ability. Men ($n$=40).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| 3:1 Quantitative performance | 3. O'Connor test<br>17. Figure classification<br>14. Block design<br>11. Reaction time | .82 | .68 | 4/35 | 18.41 | 4.7 |
| 3:2 Qualitative performance | 26. Wire labyrinth, S.D.<br>16. Leiter<br>27. Tapping, S.D.<br>1. Crawford screws | .88 | .77 | 4/35 | 29.56 | 7.6 |
| 3:3 Adjustment to work, I | 3. O'Connor test<br>6. Threading<br>9. Tapping<br>23. Identical forms | .55 | .30 | 4/35 | 3.75 | – |
| 3:4 Adjustment to work, II | 3. O'Connor test | .50 | .25 | 1/38 | 12.67 | 1.7 |
| 3:5 Total stoppage time | 5. Wire labyrinth | .41 | .17 | 1/38 | 7.54 | 1.0 |
| 3:6 Work proficiency, I | 8. MTTA<br>11. Reaction time<br>16. Leiter<br>5. Wire labyrinth<br>27. Tapping, S.D. | .93 | .87 | 5/34 | 45.42 | 12.6 |
| 3:7 Work proficiency, II | 16. Leiter<br>8. MTTA<br>25. Reaction time, S.D.<br>5. Wire labyrinth<br>27. Tapping, S.D. | .93 | .86 | 5/34 | 40.35 | 11.2 |
| 3:8 Rated quantitative performance | 16. Leiter<br>25. Reaction time, S.D.<br>26. Wire labyrinth<br>27. Tapping, S.D.<br>3. O'Connor test<br>2. Crawford pins | .86 | .74 | 6/33 | 15.47 | 4.6 |

Here also is evidenced the tendency to greater possibilities of predicting qualitative performances than quantitative performances. However, this tendency emerges only as regards one of the two criteria which comprise ratings of work proficiency.

We may summarize the results of the regression analyses by stating that it seems to be possible to predict the five performance criteria from measurements of ability with apparent success among both men and women, while, on the other hand, the possibilities of predicting the three adjustment criteria from the same variables are greatly limited.

On the whole, it is the same kind of ability variables which enter with statistically significant regression coefficients into the equations, whether the analyses refer to quantitative performances in the different tasks or to composite

*Table 47.* Prediction of eight composite criterion measures from measurements of ability. Women (*n*=

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|
| 3:1 Quantitative performance | 8. MTTA<br>19. Addition, UB<br>7. Plates | .82 | .68 | 3/20 | 14.01 | 2.8 |
| 3:2 Qualitative performance | 17. Figure classification<br>1. Crawford screws<br>9. Tapping<br>2. Crawford pins | .90 | .80 | 4/19 | 19.61 | 4.4 |
| 3:3 Adjustment to work, I | – a) | – | – | – | – | – |
| 3:4 Adjustment to work, II | 15. Raven's Matrices<br>17. Figure classification | .70 | .49 | 2/21 | 10.18 | 1.8 |
| 3:5 Total stoppage time | – a) | – | – | – | – | – |
| 3:6 Work proficiency, I | 7. Plates<br>19. Addition, UB | .76 | .57 | 2/21 | 13.92 | 2.4 |
| 3:7 Work proficiency, II | 8. MTTA<br>9. Tapping<br>14. Block design<br>3. O'Connor test<br>4. Mark-making | .88 | .77 | 5/18 | 12.17 | 2.9 |
| 3:8 Rated quantitative performance | 8. MTTA<br>19. Addition, UB<br>7. Plates<br>17. Figure classification | .85 | .72 | 4/19 | 12.02 | 2.7 |

a) No predictor variable reaches statistical significance.

*Table 48.* The correlations of the observation variables with eight composite criterion measures. Men ($n$=40).

| Predictor variables | 3:1 Quantitative performance | 3:2 Qualitative performance | 3:3 Adjustment to work,I | 3:4 Adjustment to work,II | 3:5 Total stoppage time | 3:6 Work proficiency, I | 3:7 Work proficiency, II | 3:8 Rated quantitative performance |
|---|---|---|---|---|---|---|---|---|
| 1 Emotional stability | −.34* | .09 | .43** | −.50** | .39* | .08 | −.21 | −.41** |
| 2 Ability to co-operate | −.02 | −.03 | .44** | −.36* | .37* | −.18 | .19 | −.05 |
| 3 Ability to communicate through speech | −.15 | .32 | −.25 | .05 | −.16 | .41** | −.40** | −.29 |
| 4 Motivation to work | −.48** | .28 | .48** | −.59** | .49** | .27 | −.37* | −.51** |
| 5 Movement co-ordination and working pace | −.52** | .53** | −.01 | −.31* | −.02 | .66** | −.71** | −.59** |
| 6 Technical knowledge and skills | −.68** | .63** | .20 | −.52** | .16 | .71** | −.80** | −.72** |
| 7 Rated work proficiency | .70** | −.76** | −.07 | .47** | −.03 | −.80** | .88** | .71** |
| 8 Rated adjustment to work | .60** | −.51** | −.43** | .66** | −.35* | −.40** | .48** | .52** |
| 9 Rated quantitative performance | .67** | −.49** | −.32* | .53** | −.13 | −.48** | .61** | .63** |

Note. Prediction variables 1–6 in Tables 48 and 49 have been adjusted so as to yield negative correlations with positive characteristics. ** indicates a value significant at the 1% level and * a value significant at the 5% level.

criterion measures. Psychomotor variables preponderate but are supplemented by certain spatial-inductive and numerical variables.

Only one striking difference appears and this concerns the numerical test variables in the male group. Here we find that one or the other of two numerical tests, Addition, W, and Highest number, helps to predict quantitative performances in no less than four cases, while a numerical test never enters into the equations in the corresponding analyses of composite criterion measures.

*Observational data as predictors.* If, under this heading, we first examine the zero-order correlations in the male group, we find that in Table 48 these correlations may, from a certain point of view, be said to form a logically coherent

*Table 49.* The correlations of the observation variables with eight composite criterion measures. Women ($n$=24).

| Predictor variables | 3:1 Quantitative performance | 3:2 Qualitative performance | 3:3 Adjustment to work,I | 3:4 Adjustment to work,II | 3:5 Total stoppage time | 3:6 Work proficiency, I | 3:7 Work proficiency, II | 3:8 Rated quantitative performance |
|---|---|---|---|---|---|---|---|---|
| 1 Emotional stability | −.05 | −.03 | .54** | −.41* | .54** | −.01 | −.05 | .01 |
| 2 Ability to co-operate | .02 | −.25 | .74** | −.56** | .43* | .00 | .09 | −.09 |
| 3 Ability to communicate through speech | −.38 | .38 | .02 | −.25 | .13 | .35 | −.34 | −.23 |
| 4 Motivation to work | −.05 | −.24 | .71** | −.63** | .69** | −.11 | .06 | −.20 |
| 5 Movement co-ordination and working pace | −.53** | .39 | .08 | −.27 | .13 | .62** | −.72** | −.64** |
| 6 Technical knowledge and skills | −.28 | .36 | .45* | −.62** | .33 | .37 | −.52** | −.29 |
| 7 Rated work proficiency | .59** | −.64** | −.08 | .30 | −.14 | −.68** | .83** | .59** |
| 8 Rated adjustment to work | .32 | −.04 | −.78** | .79** | −.59** | −.22 | .27 | .38 |
| 9 Rated quantitative performance | .37 | .04 | −.42* | .53** | −.41* | −.09 | .33 | .53** |

Note. Prediction variables 1--6 in Tables 48 and 49 have been adjusted so as to yield negative correlations with positive characteristics. ** indicates a value significant at the 1% level and * a value significant at the 5% level.

pattern. We may assume that both the predictor and the criterion variables are located at different points on a continuum extending from a markedly emotional reaction to a primarily cognitive or psychomotor performance, Now the position seems to be that the nearer together a predictor variable and a criterion variable lie in this continuum, the higher will be the correlation between them. Thus, the observation variables which refer to movement co-ordination and working pace, technical knowledge and skills, work proficiency and quantitative performance are throughout correlated to a significant extent with composite criterion measures which refer to quantitative or qualitative performances (3:1, 3:2, 3:6, 3:7 and 3:8). On the contrary, they are scarcely correlated at all with the adjust-

ment criteria 3:3 and 3:5. The observation variables, "ability to co-operate" and "emotional stability", on the other hand, display throughout significant correlations with all three adjustment criteria, while they are not correlated at all or only to a slight extent with the previously enumerated quantitative or qualitative performance criteria. The observation variables, "rated adjustment to work" and "motivation to work", tend to occupy an intermediate position in the assumed continuum and are also moderately correlated with most of the criteria. The observation variable "ability to communicate through speech" is significantly correlated only with criteria 3:6 and 3:7 and is consequently most closely connected with the first group of predictors.

The correlation pattern which thus appears in the male group is also to be found to a considerable extent in the female group, as an examination of Table 49 will show. The most striking feature here is that the emotionally loaded observation variables, "ability to co-operate" and "emotional stability", throughout display significant correlations with all three adjustment criteria, while they are not correlated at all with the performance criteria. Precisely the same state of affairs obtains as regards the observation variables, "motivation to work" and "rated adjustment to work", which in the female group seem to lie nearer the emotionally loaded pole in the assumed continuum. Of the four observation variables which refer more directly to cognitive or psychomotor performances, "movement co-ordination and working pace" and "rated work proficiency" behave in the same way as in the male group. Thus, they display, with one exception, significant correlations with the performance criteria and negligible correlations with the adjustment criteria. On the other hand, "technical knowledge and skills" and "rated quantitative performance" differ from the pattern, in that they are correlated to a greater extent with adjustment criteria than with performance criteria.

In all circumstances, the zero-order correlations just examined are of such an order of magnitude that there seem to be rather good possibilities of predicting composite criterion measures from observation variables. That this is really the case is also confirmed if we study the results of the stepwise multiple-regression analyses reported in Tables 50 and 51.

As regards the results in the male group, the figures in Table 50 reveal that it is possible, with apparent success, to predict performance criteria 3:1, 3:2, 3:6, 3:7 and 3:8 also from observations of behaviour, Thus, the multiple correlation coefficients vary between .76 and .89 and the corresponding $F$-values exceed the 1 % level by 4.6 times or more. We should particularly note that the first two of these criteria are objective results of measurement, while the other three are based on observations of working behaviour in the selected tasks. In spite of this difference in the form of the criteria, there is no particularly striking difference in the accuracy of the predictions, as might have been expected from Vernon's opinion,

*Table 50.* Prediction of eight composite criterion measures from observations of behaviour. Men ($n$=40).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| 3:1 Quantitative performance | 7 Rated work proficiency<br>8 Rated adjustment to work<br>9 Rated quantitative performance | .79 | .63 | 3/36 | 20.31 | 4.6 |
| 3:2 Qualitative performance | 7 Rated work proficiency | .76 | .58 | 1/38 | 52.86 | 7.2 |
| 3:3 Adjustment to work, I | 4 Motivation to work<br>2 Ability to co-operate | .59 | .34 | 2/37 | 9.70 | 1.9 |
| 3:4 Adjustment to work, II | 8 Rated adjustment to work<br>9 Rated quantitative performance<br>1 Emotional stability | .76 | .57 | 3/36 | 16.00 | 3.7 |
| 3:5 Total stoppage time | 4 Motivation to work<br>5 Movement co-ordination and working pace<br>1 Emotional stability | .63 | .39 | 3/36 | 7.79 | 1.8 |
| 3:6 Work proficiency, I | 7 Rated work proficiency | .80 | .64 | 1/38 | 68.38 | 9.3 |
| 3:7 Work proficiency, II | 7 Rated work proficiency<br>6 Technical knowledge and skills | .89 | .80 | 2/37 | 73.41 | 14.0 |
| 3:8 Rated quantitative performance | 6 Technical knowledge and skills<br>9 Rated quantitative performance | .77 | .59 | 2/37 | 26.34 | 5.0 |

that the magnitude of the correlations between predictors and criteria depends not only on the content but also on the form (Vernon 1958, Nilsson and Nordlund 1972). However, the difference which is found is in the expected direction in two cases out of three. Thus, if the results indicate that it is possible to successfully predict the performance criteria from observations of behaviour, it is at least equally noteworthy that it also seems to be possible to predict the adjustment criteria from these predictors in an acceptable manner and with greater success than could be obtained from measurements of ability. Here also there is

*Table 51.* Prediction of eight composite criterion measures from observations of behaviour. Women ($n$=24)

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|
| 3:1 Quantitative performance | 7 Rated work proficiency | .59 | .35 | 1/22 | 11.83 | 1.5 |
| 3:2 Qualitative performance | 7 Rated work proficiency<br>9 Rated quantitative performance | .74 | .54 | 2/21 | 12.41 | 2.1 |
| 3:3 Adjustment to work, I | 8 Rated adjustment to work | .78 | .61 | 1/22 | 34.34 | 4.3 |
| 3:4 Adjustment to work, II | 8 Rated adjustment to work | .79 | .62 | 1/22 | 36.02 | 4.5 |
| 3:5 Total stoppage time | 4 Motivation to work | .69 | .47 | 1/22 | 19.68 | 2.5 |
| 3:6 Work proficiency, I | 7 Rated work proficiency<br>9 Rated quantitative performance<br>8 Rated adjustment to work<br>4 Motivation to work | .83 | .69 | 4/19 | 10.55 | 2.3 |
| 3:7 Work proficiency, II | 7 Rated work proficiency | .83 | .68 | 1/22 | 47.10 | 5.9 |
| 3:8 Rated quantitative performance | 5 Movement co-ordination and working pace | .64. | .42 | 1/22 | 15.66 | 2.0 |

reason to emphasize that the form of the criteria varies, in that two of them include subjective judgements, while 3:5, "total stoppage time", is to be regarded as a more objective measurement.

A study of Table 51 reveals that the composite criterion measures can be predicted from observations of behaviour also in the female group. However, the accuracy of the predictions is comparatively small, as regards criteria 3:1 and 3:8, both of which refer to quantitative performances. More satisfactory values are to be found in the regression analyses of the qualitatively determined performance criteria 3:2, 3:6 and 3:7. However, what catches the eye, on looking at Table 51, is the fact that it seems to be possible to predict the three adjustment criteria from observations of behaviour with considerable success and better than in the male group. Finally, we may note in this connection that the accuracy of the predictions seems to be influenced to a certain extent by the form of the

criterion in the direction which Vernon presupposed. Whether we consider the quantitative and qualitative performance criteria or the three adjustment criteria, we find that the multiple correlation coefficients are somewhat higher in the cases in which the criteria are based on observations and contain judgements of a similar nature to those which occur in the predictors. However, it is worth emphasizing that the differences observed here are not particularly great. Thus, on the whole, the contents of the criteria would seem to be more decisive of the prediction possibilities than their forms.

Having analysed the possibilities of predicting composite criterion measures from observations of behaviour in both the male and the female group, we may say, by way of summary, that these possibilities are so good that it is self-evident that they should be exploited. In this connection, it seems that observational data may be a valuable complement to measurements of ability. In particular, it is evident that we must primarily rely on observational data as predictors, if we wish to predict adjustment to work. As regards the form of these observational data, the results have shown that both the sub-scales in the five-degree rating schedule and their counterparts in the general observation form of check-list type are serviceable and can supplement each other.

*Combinations of measurements of ability and observational data as predictors.* The next task will be to investigate whether the predictions of composite criterion measures can be further improved by combining measurements of ability and observations of behaviour as predictors. For this purpose, I have carried out another series of stepwise multiple-regression analyses, the results of which are shown in Tables 52 and 53. In addition, I may refer the reader to the diagrams in Figs. 26 and 27, which are intended to facilitate comparison between the efficiency of different combinations of predictors.

If we examine the given material more closely and devote our attention first to the male group, we find that the regression analyses with combinations of measurements of ability and observations of behaviour as independent variables yield results which give evidence of improved prediction possibilities. It is true that we cannot note any great gains, as regards the proportions of explained criterion variance—in relation to what was achieved with measurements of ability alone or observations of behaviour alone as predictors—but the *F*-values have increased in six cases out of eight, which means that the standard errors in the predicted values have decreased and that the accuracy of the predictions has accordingly increased. It is still true to say that the prediction of adjustment criteria presents greater difficulties than the prediction of performance criteria. However, it is noteworthy that it is possible to explain more than 50 % of the variance in the objective measurement of adjustment to work constituted by "total stoppage time". The previously noted difference between quantitatively and qualitatively determined performances also remains. That it is accordingly possible to predict

*Table 52.* Prediction of eight composite criterion measures from combinations of measurements of ability and observations of behaviour. Men ($n$=40).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| 3:1 Quantitative performance | Rated work proficiency<br>O'Connor test<br>Rated adjustment to work | .82 | .67 | 3/36 | 24.30 | 5.5 |
| 3:2 Qualitative performance | Wire labyrinth, S.D.<br>Rated work proficiency<br>O'Connor test<br>Tapping, S.D. | .89 | .79 | 4/35 | 32.43 | 8.3 |
| 3:3 Adjustment to work, I | Motivation to work<br>Ability to co-operate | .59 | .34 | 2/37 | 9.70 | 1.9 |
| 3:4 Adjustment to work, II | Rated adjustment to work<br>O'Connor test | .73 | .53 | 2/37 | 20.91 | 4.0 |
| 3:5 Total stoppage time | Motivation to work<br>Movement co-ordination and working pace<br>Crawford screws<br>Tapping, S.D. | .71 | .51 | 4/35 | 9.09 | 2.3 |
| 3:6 Work proficiency, I | MTTA<br>Reaction time<br>Leiter<br>Wire labyrinth<br>Tapping, S.D. | .93 | .87 | 5/34 | 45.42 | 12.6 |
| 3:7 Work proficiency, II | Rated work proficiency<br>O'Connor test<br>Figure classification<br>Reaction time | .95 | .89 | 4/35 | 73.94 | 18.9 |
| 3:8 Rated quantitative performance | Technical knowledge and skills<br>Leiter<br>Rated quantitative performance | .81 | .66 | 3/36 | 23.29 | 5.3 |

the latter with somewhat greater success would seem to be connected with the fact that they are criterion variables, which are more homogeneous than the former.

In the female group also, the results indicate that the prediction possibilities are, as a rule, better when measurements of ability and observations of behaviour

*Table 53.* Prediction of eight composite criterion measures from combinations of measurements of ability and observations of behaviour. Women ($n$ = 24).

| Criteria | Predictors | Multiple correlation | Prop. explained variance | No. of Df | $F$-value | No. of times $F$-value exceeds 1% point |
|---|---|---|---|---|---|---|
| 3:1 Quantitative performance | MTTA<br>Ability to communicate through speech<br>Motivation to work<br>Plates | .87 | .75 | 4/19 | 14.35 | 3.2 |
| 3:2 Qualitative performance | Figure classification<br>Rated work proficiency<br>Crawford screws<br>Ability to communicate through speech | .91 | .83 | 4/19 | 22.52 | 5.0 |
| 3:3 Adjustment to work, I | Rated adjustment to work | .78 | .61 | 1/22 | 34.34 | 4.3 |
| 3:4 Adjustment to work, II | Rated adjustment to work<br>Raven's matrices | .84 | .70 | 2/21 | 24.89 | 4.3 |
| 3:5 Total stoppage time | Motivation to work<br>Mark-making | .77 | .60 | 2/21 | 15.58 | 2.7 |
| 3:6 Work proficiency, I | Rated work proficiency<br>Plates<br>Block design | .83 | .69 | 3/20 | 14.66 | 3.0 |
| 3:7 Work profiency, II | Rated work proficiency<br>Plates | .89 | .79 | 2/21 | 39.16 | 6.8 |
| 3:8 Rated quantitative performance | Movement co-ordination and working pace<br>Addition, UB<br>Plates | .79 | .63 | 3/20 | 11.39 | 2.3 |

are combined as predictors. The proportions of explained variance in the different criteria vary between .60 and .83, and the corresponding $F$-values exceed the 1 % level by between 2.7 and 6.8 times. These are highly satisfactory values. We may especially note that the adjustment criteria can also be successfully predicted. The difference noted in the male group between quantitative and qualitative performances, which meant that the prediction possibilities were better as regarded the latter, also appears in the female group, though in a somewhat less marked fashion.

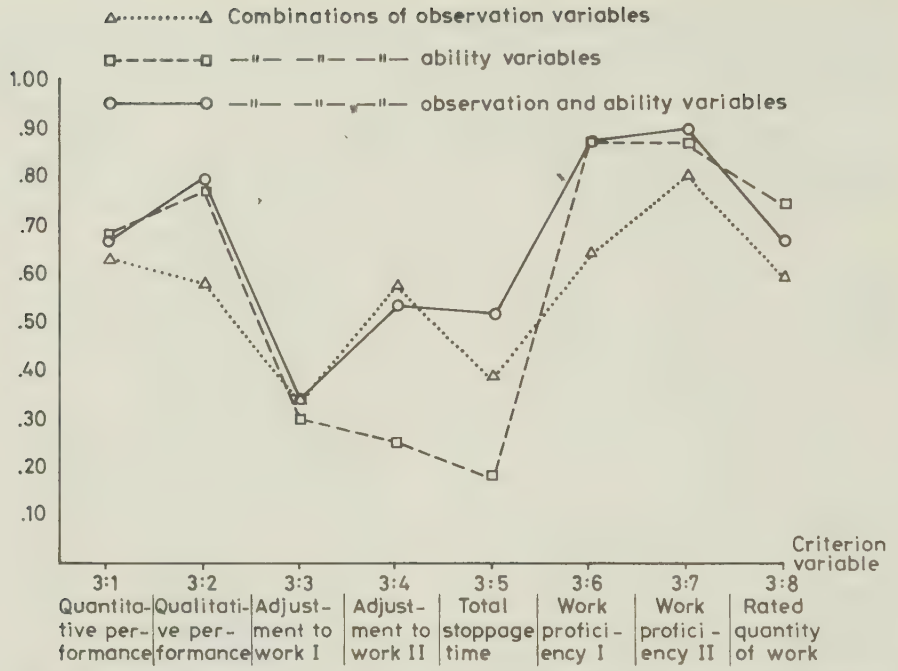



Fig. 26. Proportions of explained criterion variance in the prediction of eight composite criterion measures from different combinations of predictors. Men ($n$=40).

## C. Cross-validation studies

There was no possibility of cross-validating the reported results by parallel investigations of another sample, nor is such an opportunity likely to be offered in the foreseeable future. I therefore had to have recourse to the method which I had already used for complementary purposes as regarded the prognosis-group and percentage-group criteria, and which is designated double cross-validation or sample fractionation. While this procedure is associated with certain disadvantages, once more definitive results are sought, it can be used in any case to get "a general idea of the extent of shrinkage" (Horst 1966). As shown in the foregoing pages, the total experimental group was reduced in this part of the investigation from 80 to 64, so that it is not feasible, in dividing up the group, to keep men and women separate, however desirable this might be. Under such circumstances it seemed natural to divide up the experimental group by assigning working groups nos. 1 and 3 to the one sub-sample and working groups nos. 2 and 4 to the other (cf. above, p. 62). This means that each sub-sample consisted of altogether 20 men and 12 women.

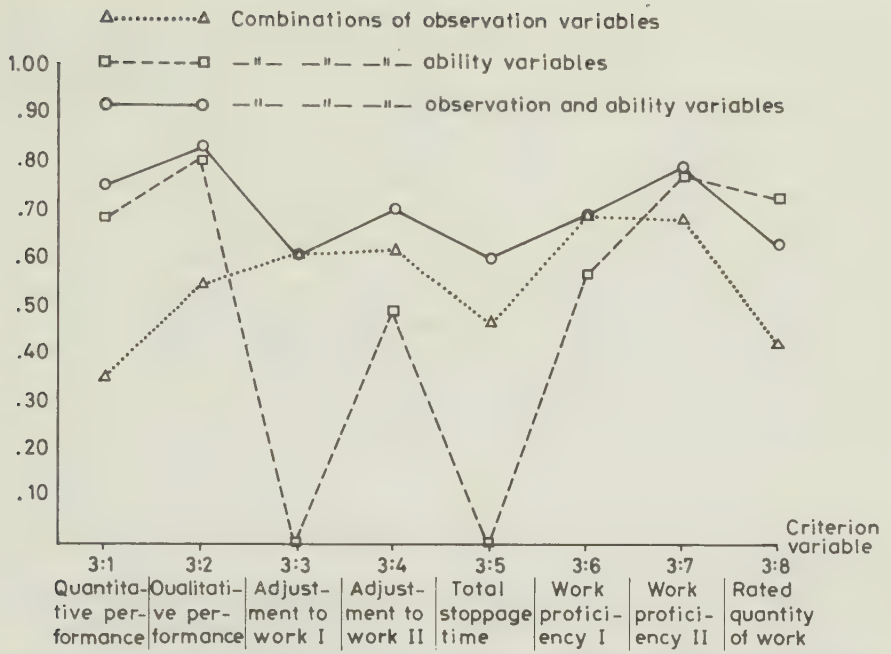



Fig. 27. Proportions of explained criterion variance in the prediction of eight composite criterion measures from different combinations of predictors. Women ($n$=24).

It would be going too far—and cannot be regarded as necessary—to carry out cross-validations of all criteria. For this reason, I selected from them three variables which together measure the three main aspects of working capacity which I previously distinguished. These variables are "Average quantitative performance", "Average qualitative performance" and a composite measure of adjustment to work. The latter calls for special comment.

In the foregoing pages we studied three different measures of adjustment to work, namely, 3:3, "Average rating of adjustment to work, I", 3:4, "Average rating of adjustment to work II", and 3:5 "Total stoppage time". These display moderate or high inter correlations and were assigned to one and the same factor in a factor analysis (cf. above p. 95). In order to obtain a more comprehensive measure of adjustment to work than that constituted by "Total stoppage time", for example, I applied the same procedure as in the case of quantitative performance and thus converted individual raw scores in the three variables into equivalent $z$ scores and calculated their averages. The values calculated in this way were then expressed on a scale, with 50 as the mean value and 10 as the standard deviation and display the distribution shown in Fig. 28. Low values denote satisfactory adjustment to work. Bearing this in mind we may see in Fig. 28 what was found when the three scales were analysed separately, namely, that

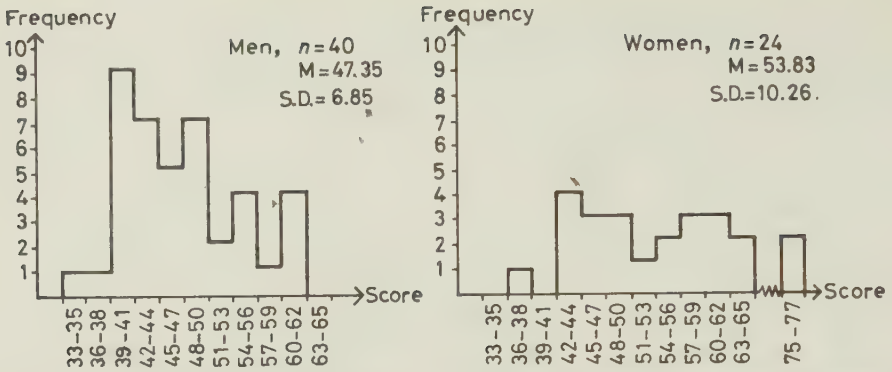



Fig. 28. Distributions of composite measures of adjustment to work (average of the *z*-values of three sub-scales) among mentally retarded men and women.

the men show better adjustment to work than the women. The difference is statistically significant at the 1 % level.

To test the accuracy with which this more complex measure of adjustment to work can be predicted, I carried out stepwise multiple-regression analyses with combinations of measurements of ability and observations of behaviour as predictors among both men and women. The results are shown in Table 54.

Here we find that the observations of behaviour are effective predictors of this composite measure of adjustment to work and that the accuracy of the prediction is greater among women than among men. This agrees in both respects with my previous findings, when I analysed the possibilities of predicting the three scales, each of which measures adjustment to work.

After this digression on measures of adjustment to work I shall now return to my account of the cross-validation studies. I began these by performing stepwise

*Table 54.* Prediction of composite measures of adjustment to work from combinations of measurements of ability and observations of behaviour in men and women.

| Sample | Predictors | Multiple correlation | Prop. explained variance | No. of Df | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Men (*n*=40) | Motivation to work<br>Ability to co-operate | .70 | .49 | 2/37 | 17.49 | 3.3 |
| Women (*n*=24) | Rated adjustment to work<br>Motivation to work<br>Block design | .89 | .77 | 3/20 | 24.62 | 5.0 |

*Table 55.* Prediction of three criteria from combinations of measurements of ability and observations of behaviour in two sub-samples.

| Criteria and sub-samples | Predictors | Multiple correlation | Prop. explained variance | No. of Df | *F*-value | No. of times *F*-value exceeds 1% point |
|---|---|---|---|---|---|---|
| Average quantitative performance Sub-sample 1 | Rated work proficiency<br>Raven's Matrices<br>Rated adjustment to work | .84 | .71 | 3/28 | 22.79 | 5.0 |
| Average quantitative performance Sub-sample 2 | Wire labyrinth<br>Rated work proficiency | .75 | .56 | 2/29 | 16.68 | 3.4 |
| Average qualitative performance Sub-sample 1 | Rated work proficiency<br>Wire labyrinth, S.D.<br>Figure classification | .83 | .69 | 3/28 | 20.65 | 4.5 |
| Average qualitative performance Sub-sample 2 | Rated work proficiency<br>Crawford screws<br>Figure classification<br>Movement co-ordination and working pace | .91 | .82 | 4/27 | 31.48 | 7.7 |
| Composite measure of adjustment to work Sub-sample 1 | Rated adjustment to work | .77 | .60 | 1/30 | 45.00 | 6.0 |
| Composite measure of adjustment to work Sub-sample 2 | Motivation to work<br>Emotional stability<br>Rated quantitative performance<br>Movement co-ordination and working pace<br>Ability to communicate through speech | .89 | .79 | 5/26 | 20.08 | 5.3 |

multiple-regression analyses of the selected criteria in the two sub-samples. The results of these analyses are given in Table 55.

If we study the values in the table, we find that these bear witness to good prediction possibilities, despite the fact that both men and women are included in both sub-samples. The composition of predictors is more or less that expected but varies to some extent from the one sub-sample to the other, which is a natural consequence of the fact that several variables display approximately the same high correlations with the criteria.

After calculating the predicted criterion values, I correlated these with the

*Table 56.* Cross-validation coefficients referring to three selected criteria in two sub-samples.

| Average quantitative performance | | Average qualitative performance | | Composite measure of adjustment to work | |
|---|---|---|---|---|---|
| Sub-sample 1 $n$=32 | Sub-sample 2 $n$=32 | Sub-sample 1 $n$=32 | Sub-sample 2 $n$=32 | Sub-sample 1 $n$=32 | Sub-sample 2 $n$=32 |
| .64 | .61 | .82 | .76 | .54 | .72 |

observed values in each group and thus obtained the cross-validation coefficients given in Table 56. These are unusually high, especially as regards "Average qualitative performance". Thus, the extent of shrinkage is moderate and the results of the cross-validation provide convincing confirmation that different aspects of working capacity among the mentally retarded can be successfully predicted from conbinations of measurements of ability and observations of behaviour.

## 3. The Predictor Structure of the Ability Variables

**A. The results of factor analyses**

In Chapter 3, I gave a detailed account of the selection of prediction variables. This account makes it clear that for various reasons I found it justifiable to analyse the prediction value partly in measurements of ability in the psychomotor, spatial-inductive and numerical spheres and partly in personality characteristics, as they can be observed in ordinary working situations. In the selection of measurements of ability, I used predominantly tests which had previously been subjected to factor analysis, either by Kebbon or by other researchers. On the other hand, I made no factor analyses of my own, in view of the facts that the number of variables was large and the number of experimental subjects comparatively small. On this account, the classification in Chapter 3 of the tests used from the factor point of view should be regarded as essentially preliminary. In connection with the recurrent examinations of the prediction results, I also felt an increasing need to be able to express with greater certainty my views on the predictor structure. In order to be able to do so at a later stage, I proceeded as follows.

I first made a collation of all the tests which had entered as independent variables with significant regression coefficients into any of the previously obtained regression equations and which had also been included in the investigations of both the main sample and the cross-validation sample in southern Sweden. The number of these tests proved to be 21. The main sample and the cross-validation sample were then combined. The total number of men was

*Table 57.* Factor matrix for 21 predictor variables among men ($n$=87).

| Predictor variables | Rotated factor loadings | | | | | | $h^2$ |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1. Crawford screws | .12 | −.22 | .20 | −.27 | .70 | −.17 | .69 |
| 2. Crawford pins | .15 | −.28 | .23 | −.43 | .61 | .00 | .71 |
| 3. O'Connor test | .21 | −.24 | .16 | −.40 | .56 | .03 | .59 |
| 4. Mark-making | .47 | −.44 | .18 | −.25 | .24 | .16 | .59 |
| 5. Wire labyrinth | −.13 | .77 | −.06 | .23 | −.37 | .29 | .89 |
| 6. Threading | −.27 | .51 | −.06 | .07 | −.58 | −.10 | .69 |
| 8. MTTA | −.15 | .58 | −.12 | .49 | −.37 | .01 | .74 |
| 9. Tapping | −.27 | .81 | −.12 | .05 | −.19 | −.03 | .77 |
| 10. Copying | .30 | −.36 | .04 | −.71 | .17 | −.07 | .77 |
| 11. Reaction time | −.19 | .16 | −.82 | .20 | −.17 | .03 | .80 |
| 13. Surface development | .15 | −.06 | .14 | −.70 | .25 | −.07 | .61 |
| 14. Block design | .38 | −.04 | .20 | −.67 | .20 | −.13 | .70 |
| 15. Raven's Matrices | .37 | −.03 | .14 | −.72 | .09 | −.17 | .71 |
| 19. Addition, UB | .85 | −.08 | .09 | −.23 | .13 | −.14 | .82 |
| 20. Addition, W | .74 | −.27 | .26 | −.33 | .07 | .10 | .81 |
| 22. Highest number | .80 | −.09 | .22 | −.21 | .22 | −.03 | .79 |
| 23. Identical forms | .61 | −.25 | .08 | −.48 | .15 | .07 | .71 |
| 24. Vocabulary | .33 | −.22 | .08 | −.47 | .24 | .24 | .50 |
| 25. Reaction time, S.D. | −.24 | .11 | −.83 | .15 | −.17 | .00 | .81 |
| 26. Wire labyrinth, S.D. | −.04 | .58 | −.09 | .42 | −.27 | .46 | .81 |
| 27. Tapping, S.D. | −.06 | .78 | −.12 | .04 | −.05 | −.06 | .63 |
| Total contribution of variance | 3.36 | 3.53 | 1.79 | 3.65 | 2.30 | .51 | 15.14 |

thereby 87 and the total number of women 55. In these combined groups, correlations were calculated between the 21 tests, after which these correlations were finally subjected to factor analysis. In this case also, the number of variables is, of course, large in relation to the number of individuals, particularly in the female group. Nevertheless, I considered it justifiable and meaningful to carry out the analyses, partly in order to find out to what extent the results of the previously reported factor analyses would be confirmed and partly in order to see whether any further factor differentiation would appear in the quite large number of psychomotor variables.

The factor analyses were carried out by Jöreskog's J method, using the FAIM programme. In the male group, the $T_K$ values and the significance test indicate that we should reckon upon six factors to explain the common variance. Though the last of these six factors is responsible for only a small contribution of variance and can scarcely be interpreted in an unexceptionable way, I found it reasonable to include it, as this would increase the possibilities of making comparisons with the female group. In this group, the $T_K$ values indicate that we should reckon

*Table 58.* Factor matrix for 21 predictor variables among women ($n$=55).

| Predictor variables | Rotated factor loadings | | | | | | $h^2$ |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1. Crawford screws | .06 | .33 | .15 | .05 | −.43 | .51 | .59 |
| 2. Crawford pins | .31 | .14 | −.02 | .27 | −.33 | .52 | .57 |
| 3. O'Connor test | .19 | .32 | .15 | .01 | −.14 | .78 | .79 |
| 4. Mark-making | .24 | .10 | .50 | .02 | −.25 | .23 | .44 |
| 5. Wire labyrinth | −.09 | −.80 | −.09 | −.08 | .39 | −.22 | .85 |
| 6. Threading | −.19 | −.31 | −.36 | .03 | .37 | −.44 | .59 |
| 8. MTTA | −.19 | −.22 | −.17 | −.19 | .73 | −.26 | .76 |
| 9. Tapping | −.43 | −.17 | −.34 | −.22 | .47 | −.10 | .61 |
| 10. Copying | .14 | .19 | .12 | .49 | −.63 | .16 | .74 |
| 11. Reaction time | −.84 | −.06 | −.24 | −.10 | .09 | −.17 | .81 |
| 13. Surface development | −.08 | −.08 | .10 | .57 | −.17 | .11 | .38 |
| 14. Block design | .13 | .27 | .23 | .61 | −.18 | .12 | .56 |
| 15. Raven's Matrices | .21 | .26 | .12 | .75 | −.06 | −.06 | .70 |
| 19. Addition, UB | .12 | .06 | .79 | .13 | −.10 | .11 | .68 |
| 20. Addition, W | .21 | −.02 | .79 | .45 | −.07 | .13 | .90 |
| 22. Highest number | .13 | −.06 | .73 | .31 | −.07 | −.08 | .65 |
| 23. Identical forms | .21 | .07 | .28 | .48 | −.05 | −.05 | .37 |
| 24. Vocabulary | .17 | −.05 | .41 | .52 | .03 | .42 | .65 |
| 25. Reaction time, S.D. | −.92 | .00 | .18 | .15 | .09 | −.10 | .92 |
| 26. Wire labyrinth, S.D. | −.04 | −.80 | .10 | −.25 | .09 | −.24 | .77 |
| 27. Tapping, S.D. | −.83 | −.11 | −.12 | −.15 | .18 | −.16 | .79 |
| Total contribution of variance | 2.93 | 1.89 | 2.83 | 2.62 | 1.95 | 1.90 | 14.12 |

upon seven factors, while the significance test indicates only five. However, such circumstances as the amounts of the contributions of variance and the interpretability of the factors justifies our studying primarily the six-factor solution in this group also. With reference to these considerations, I give in Tables 57 and 58 the rotated factor matrices and the communality values for the 21 predictor variables among men and women.

A closer study of the factor loadings in the male group provides us with a basis for a psychological interpretation of the first five factors in the following manner. Factor 1 measures numerical ability and has high loadings in the two addition tests, as well as in "Highest number". There are grounds for also assigning "Identical forms" and "Mark-making" to this factor. Factor 2 measures psychomotor ability, as it appears in "Tapping", "Tapping, S.D.", "Wire labyrinth", "Wire labyrinth, S.D." and MTTA. Factor 3 measures rapidity and consistency in simple reactions and has high loadings in "Reaction time" and "Reaction time, S.D.". Factor 4 measures spatial-inductive ability and has high loadings in "Raven's Matrices", "Copying", "Surface development" and "Block design".

"Vocabulary" also should primarily be assigned to this factor, which may seem surprising to begin with but which becomes comprehensible if we recall that the vocabularies of the experimental subjects were tested by giving them the task of stating what a number of figure drawings represented. Factor 5 measures psychomotor ability, as it appears in "Crawford screws", "Crawford pins", "Threading" and the O'Connor Test. Factor 6, finally, is characterized throughout by low loadings in the different variables. "Wire labyrinth, S.D." is a certain exception in this respect; however, the conclusion with regard to this factor will be that we do not have a sufficient basis for a psychological interpretation of it.

If we then pass on to study the factor matrix in the female group, we arrive at the following interpretations of the six factors. Factor 1 has high loadings in the two reaction-time measurements and also in "Tapping, S.D." and—though it is less marked—in "Tapping". It may be asserted that this factor measures rapidity and ability to concentrate in simple psychomotor reactions. Factor 2, with high loadings in "Wire labyrinth" and "Wire labyrinth, S.D.", measures manual ability in a task which also demands concentration and persistence. Factor 3 is a counterpart to Factor 1 in the male group and thus measures numerical ability, with high loadings in the two addition tests and "Highest number" and also—though it is less marked—in "Mark-making". Factor 4 corresponds to Factor 4 in the male group. It measures, above all, spatial-inductive ability and has its highest loadings in "Raven's Matrices", "Block design" and "Surface development". As in the male group, there are grounds for also assigning "Vocabulary" to this factor. Factor 5 measures a certain aspect of psychomotor ability. It has its highest loadings in MTTA and "Copying" and may therefore merit the designation of "the tracing and copying factor". Factor 6, finally, corresponds closely to Factor 5 in the male group and thus also measures a certain kind of psychomotor ability.

These factor interpretations show clearly that there is obvious agreement between the factor patterns in men and women. Thus, in both groups we find clear evidence of a numerical factor, a spatial-inductive factor and a psychomotor factor—the latter has comparatively high loadings in "Crawford screws", "Crawford pins", the O'Connor Test and "Threading". In all these respects, we may also note that the results confirm what Kebbon found in his more extensive factor-analytical investigations of mentally retarded men and women (Kebbon 1965, Kebbon, Aldrin and Hallerstedt 1967).

The agreement with Kebbon's results—even in details—is remarkable and the circumstance that the average intelligence level was lower in my experimental groups does not seem to have been of any decisive importance, as regards the result of the analysis. It is true that there seems to be a certain amount of hesitation, as regards the factor affiliation of one of the tests included in both Kebbon's

and my own studies, namely, "Mark-making". I found reason to assign this test to a numerical factor, while Kebbon seems to regard it as predominantly psychomotor in character. If we examine his different factor matrices in detail, however, it turns out that, in fact, "Mark-making" displays a higher loading in the numerical factor than in the psychomotor factor in his cross-validation sample. For this reason, the correct conclusion would seem to be that this test is not particularly serviceable or suitable as a factor test. For example, it may very well be that, as a consequence of changes in the subjects' attitudes to the task, the test may sometimes measure especially a psychomotor factor, while in another connection it may measure especially a numerical factor.

An analogous conclusion seems to be justifiable, as regards "Identical forms". Kebbon regards this test as predominantly numerical in character, and the result in the male group which I have just reported supports this view. However, here the test shows a comparatively high loading also in the spatial-inductive factor and, if we study the result of the analysis in the female group, we find that here "Identical forms" involves primarily the spatial-inductive factor. Precisely the same result is also to be found in Kebbon's cross-validation sample. For this reason, "Identical forms" also does not seem to be particularly serviceable as a factor test in studies of mentally retarded persons.

As I am unable in this connection to make any real contribution to the interesting discussion on ability structures at different levels of ability, I would only add here that we may, of course, also regard observations of overlapping between different factors as signs that in certain respects the ability structure seems to be less differentiated in the mentally retarded than in persons of normal intelligence (cf. Kebbon 1965 and Garret 1946).

However, the similarities between the factor patterns in men and women are not restricted to what has already been demonstrated, viz. that in both groups a numerical, a spatial-inductive and a psychomotor factor are to be found with the same contents, but are also expressed in other ways. Thus, in the male group, Factor 3 is a reaction-time factor which occurs almost unchanged as Factor 1 in the female group. However, among the women, the scope of this factor seems to be somewhat wider, in that "Tapping, S.D." also displays a high loading in it. That this should be the case is not really surprising, if we recall the actual test situations. In order to give proof of a high intra-subject variability in "Tapping, S.D.", it is necessary to maintain greater watchfulness during the whole test procedure. One may very well imagine that this watchfulness will be of somewhat the same kind as that which is prerequisite for satisfactory performances in "Reaction time" and "Reaction time, S.D.". However, among the men, no such relation is to be found between "Tapping, S.D." and the reaction-time measurements. This may be interpreted as a sign that the behaviours triggered off by the test situations were not characterized by exactly the same organization in men and women.

One reaches the same conclusion when one compares the remaining factors (Nos. 2 and 6) in the male group with the corresponding factors (Nos. 2 and 5) in the female group. Among the men, Factor 2 is a wide psychomotor factor, which, together with the intrinsically inconsiderable Factor 6, is responsible for 27 % of the common variance. To definitely distinguish Factor 2 from Factor 5, which is also a factor in the sphere of psychomotor function, presents certain difficulties with the information available at present. In both cases, manual ability and rapidity are required. It is scarcely possible to demonstrate any consistent differences, as regards the character of the movement sequences or the method of evaluation. On the other hand, it may very well be that the tests included in Factor 2—"Tapping", "Wire labyrinth" and MTTA—are slightly frustrating and thereby make greater demands on the ability to concentrate and on persistence.

In the female group, Factors 2 and 5 measure largely the same psychomotor ability as Factors 2 and 6 in the male group. Among the women, however, we also find here signs of a somewhat different behavioural organization, which is expressed by the fact that the two factors are of approximately equal strength, in the sense that they contribute 13 % and 14 % to the common variance respectively. Factor 2 had high loadings in "Wire labyrinth" and "Wire labyrinth, S.D.". These tests are thus distinguished from MTTA, "Copying" and "Tapping", which are the tests which have high loadings in Factor 5.

The results of the factor analyses just commented on may be summarized in the way shown in Tables 59:1 and 60:1. In these tables, all 21 tests have been assigned to the respective factors. Moreover, the predictor variables entitled "Leiter Scale", "Figure classification" and "Plates" have been classified and entered in the appropriate columns. These three tests were not included in the studies of the cross-validation sample and therefore could not be included in the factor analyses reported here. However, a study of the intercorrelations, as well as of the reports of factor analyses performed in other connections, shows that "Leiter Scale" and "Figure classification" measure spatial-inductive ability and should accordingly be assigned to a factor with this name. In the same way, there are grounds for assigning "Plates" to "Psychomotor factor II".

In connection with the interpretation of the factors, I have noticed that a form of experimental dependence may be present among, on the one hand, the variables "Reaction time", "Wire labyrinth" and "Tapping" and, on the other hand, the three corresponding measures of intra-subject variability. In order to throw light upon the question if this circumstance has in any decisive sense distorted the results, I have repeated the analyses without including the variability measures. In the male group, then, the reaction-time factor dropped out, as expected, but in all other respects the same ability structure appeared as before, with one numerical factor, one spatial-inductive factor and two psychomotor factors. The magnitude of their contribution to the variance changed only slightly.

*Table 59:1.* Ability structure among men ($n$=87).

| Factor designation | Numerical factor | Psychomotor factor I | Psychomotor factor II | Reaction-time factor | Spatial-inductive factor |
|---|---|---|---|---|---|
| Ability variables belonging to a certain factor | Addition, UB<br>Highest number<br>Addition, W<br>Identical forms<br>Mark-making | Tapping<br>Tapping, S.D.<br>Wire labyrinth<br>Wire labyrinth, S.D.<br>MTTA | Crawford screws<br>Crawford pins<br>Threading<br>O'Connor test | Reaction-time<br>Reaction time, S.D. | Raven's Matrices<br>Copying<br>Block design<br>Vocabulary<br>The Leiter scale<br>Figure classification |

*Table 59:2.* The number of times which ability variables belonging to a certain factor entered into regression equations obtained in altogether 22 stepwise multiple-regression analyses of the three different criterion groups with statistically significant regression coefficients.

| | Numerical factor | Psychomotor factor I | Psychomotor factor II | Reaction-time factor | Spatial-inductive factor |
|---|---|---|---|---|---|
| Six analyses of prognosis-group and percentage-group criteria | 2 | 3 | 5 | 3 | 11 |
| Eight analyses of quantitative performances in selected tasks | 5 | 5 | 4 | 1 | 2 |
| Eight analyses of composite criterion measures | 1 | 12 | 7 | 4 | 6 |

*Table 60:1.* Ability structure among women ($n$=55).

| Factor designation | Numerical factor | Psychomotor factor I | | Psychomotor factor II | Reaction-time factor | Spatial-inductive factor |
|---|---|---|---|---|---|---|
| | | I:1 | I:2 | | | |
| Ability variables belonging to a certain factor | Addition, W<br>Addition, UB<br>Highest number<br>Mark-making | Wire labyrinth<br>Wire labyrinth, S.D. | MTTA<br>Copying<br>Tapping | O'Connor test<br>Crawford pins<br>Crawford screws<br>Threading<br>Plates | Reaction time, S.D.<br>Reaction time<br>Tapping, S.D. | Raven's Matrices<br>Block design<br>Surface development<br>Vocabulary<br>Identical forms<br>Figure classification |

*Table 60:2.* The number of times which ability variables belonging to a certain factor entered into regression equations obtained in altogether 18 stepwise multiple-regression analyses of the three different criterion groups with statistically significant regression coefficients.

| | Numerical factor | Psychomotor factor I:1 | Psychomotor factor I:2 | Psychomotor factor II | Spatial-inductive factor |
|---|---|---|---|---|---|
| Two analyses of prognosis-group and percentage-group criteria | – | – | 2 | 2 | 1 |
| Eight analyses of quantitative performances in selected tasks | 4 | 1 | 6 | 5 | 2 |
| Eight analyses of composite criterion measures | 4 | – | 5 | 6 | 5 |

In the female group the reaction time factor also dropped out and as well the factor which is called "Psychomotor factor I:1" in Table 60:1 and which contains high loadings in the variables "Wire labyrinth" and "Wire labyrinth, S.D.". The displacements of the contributions to the variance of the remaining factors are somewhat larger here, but the total picture is unchanged. Thus, in the female group too, one numerical factor, one spatial-inductive factor and two psychomotor factors still appear. Therefore, there is ground for the conclusion that the suspected experimental dependence among some of the ability variables has not, in any decisive sense, distorted the results of the factor analyses.

**B. The predictive power of the different ability groups**

After this analysis of the ability structure, let us once more return to the question of the relative importance of the different predictor groups. The figures in Tables 59:2 and 60:2 express the frequency with which ability variables belonging to a certain factor have entered into regression equations obtained in stepwise multiple-regression analyses of different criterion groups with statistically significant regression coefficients. Though they are not exact measurements of the prediction value, these figures give a certain idea of the relative importance of the factors included in the predictor structure. Since none of the variables which belong to the reaction-time factor contribute to the prediction of the criteria in the female group, this factor could be omitted in Table 60:2. In spite of this, however, the number of columns here is unchanged in relation to Table 59:2, which is a consequence of the fact that, on the other hand, we have reason to distinguish two subsidiary factors in the female group under the designation "Psychomotor factor I".

At the beginning of Chapter V, it was possible to demonstrate, after making special studies concerning the prediction of the prognosis-group and percentage-group criteria, that tests in the sphere of psychomotor function are generally the best prediction variables when we wish to predict working capacity in the mentally retarded. This conclusion is not contradicted by a study of Tables 59:2 and 60:2. Here the predictive power of psychomotor-ability variables seems to emerge even more strongly when the criteria consist of quantitative performances in the eight tasks or of composite criterion measures, at any rate in the male group. On the other hand, these frequency figures also confirm that the accuracy of the predictions is improved in several cases by including spatial-inductive and numerical variables in the predictor combinations.

# Chapter VI

# Summarizing Discussion and Conclusions

Having reported in the preceding chapter the main results of my prediction investigations, it is now time for me to give a retrospective summary, in order to show the extent to which the proposed aims of the investigations could be achieved and to discuss the conclusions to which the results give rise in other respects.

## 1. Summarizing Discussion

*The starting-points for the investigations*

In these investigations, I started with the realization, to which people in various countries are slowly awakening, that society has a responsibility to care for the mentally retarded, not only when they are children but also when they become adults. A long series of follow-up investigations has shown that the pessimistic attitude which formerly prevailed regarding the prospects for the occupational and social adjustment of the mentally retarded is to a large extent unjustified. The need of individual guidance, aiming at occupational training and future adjustment to work, is, however, remarkably great. If such individual guidance is to be effective and successful, it must be based on reliable prognostic information. It is the realization of this which has urged on the numerous prediction investigations which have seen the ligtht of day in this field.

However, the results of these prediction investigations have to a large extent been contradictory and several authors have pointed out numerous methodological deficiencies in them. In a recently published survey, the author speaks, with a certain amount of justification, of "the monumental failure we have experienced in trying to develop valid predictors" and recommends unhesitatingly that "research on evaluation should move away from prediction" (Gold 1973). In recent years, however, certain advances have been made which seem to have been principally connected with two circumstances. On the one hand, attempts have been made to combine several different variables determined by methods of differential psychology as predictors and, on the other, the problems associated with criteria have begun to be analysed in a more purposeful manner than before. By these means, some promising results have been achieved,

but successful cross-validations are still lacking and replications of the investigations made are few and far between.

My own investigations were connected with the above-mentioned lines of methodological development and I was able, in an unusually favourable milieu for research, to observe and measure a large number of prediction variables, to analyse in an empirical sense the structure of the criterion variable and to measure various aspects of it in realistic testing programmes, which were closely co-ordinated with the regular work in a sheltered workshop.

*Selection and measurement of prediction variables*

As far as the selection of predictors was concerned, I came to the conclusion, from an examination of previous investigations and other considerations, that I should combine ability variables—chiefly in the psychomotor field but also in the spatial-inductive and numerical fields—and personality variables; the latter could be measured by systematic observations of the behaviour in different types of work situation. In Chapter III, I have given a detailed account of the measurements made of 27 different ability variables. The reliability of these measurements was found to be, on the whole, clearly satisfactory, while, on the other hand, certain shortcomings appeared as regarded their ability to differentiate. However, after omitting variables with extremely skewed distributions, I decided that these shortcomings were tolerable and were primarily due to the fact that the experimental subjects were a highly selected group. Certain differences between the sexes could be observed, both as regarded the average level of performance and as regarded the intercorrelation pattern, and were considered to justify men and women being treated as two different groups in the subsequent analysis of the data. In the same chapter, I also showed how systematic observations could be made and could result in the design of nine different prediction variables with contents extending from purely emotional to purely cognitive behaviour in working situations. While these variables also displayed more or less strong tendencies to skewed distributions, they were found to be almost as reliable as the ability variables, as was later confirmed by the calculation of both coefficient alphas and the intra-class correlations.

*Analysis of the content of the criterion variable*

As regarded the content of the criterion variable, my preparatory studies led me to the view that I should refrain from attemting to make comprehensive measurements of social adjustment, which is a particularly complex phenomenon. Instead, the criterion variable should be restricted to refer to work performances and adjustment to work. By this means, it might be possible to avoid too difficult measuring problems, while concentrating on factors that were highly relevant from the point of view of habilitation.

On this basis, I defined the criterion variable as a primarily quantitatively determined ability to carry out approved performances at work in a sheltered workshop for the mentally retarded, and during the first stage of my investigations I procured three ratings of this ability, viz. the prognosis-group criterion, the percentage-group criterion and the wages-group criterion. As a result of a certain loss in the last-mentioned variable, I used only the prognosis-group criterion and the percentage-group criterion in my further studies. These two criteria display comparatively high intercorrelations but are not for that reason identical in content. By analysing their correlations with other criterion variables, I was able to show it to be probable that the prognosis-group criterion takes greater account of qualitative skill and developmental capacity in the individuals, while the percentage-group criterion reflects to a greater extent the actual capability from the quantitative point of view. Both are based on observations and ratings made by a number of supervisors trained in observation methods. The calculated intra-class correlations show that both the prognosis-group criterion and the percentage-group criterion fulfil high requirements as regards reliability.

*Measurements of work performances*

However, right from the beginning of the investigations, I intended to make direct measurements of work performances, in order thereby to obtain fully objective criteria, and in Chapter IV a detailed account will be found of how these intentions were realized in a later stage. A test programme, including eight tasks, was organized. Four of these tasks were regular piece-work tasks at the ALA Foundation's workshop, while the rest were specially designed to serve only as experimental tasks. The participants in the experiment had to work for a total of four days at each of these tasks and during this period the progress of their performance was analysed in detail from both the qualitative and the quantitative points of view. Measurements of the quantitative performances in the different tasks were finally calculated as the number of approved products produced per hour during the last two working days taken together.

Since the work performances were chosen so as to be representative of the productive work done at well-organized sheltered workshops for the mentally retarded, it may be maintained that the quantitative measurements calculated in this way are good criteria of working capacity. I was able to show that they fulfil very high requirements as regards reliability. Moreover, their intercorrelations have been examined and factor-analysed. The results indicate that there are noticeable differences, as regards the intercorrelation pattern in men and women, but that it is nevertheless possible to distinguish two factors of the same kind in both groups.

*Observations of working behaviour and composite criterion variables*

Besides the above-mentioned criteria, I also developed and studied a number of composite criterion variables, which are based on the measurements and systematic observations made in connection with all eight tasks. The inter-correlations of these variables were also examined and factor-analysed. The results obtained show that among these composite criterion variables it is possible to distinguish three different groups, each of which measures its own aspect of the working behaviour—the quantitative aspect, the qualitative aspect and, finally, an aspect which has reference to motivation and adjustment in working situations.

*Prediction of prognosis-group and percentage-group criteria*

My first real prediction investigations involved studying the possibilities of predicting the prognosis-group and the percentage-group criteria from different combinations of ability variables. As means in my endeavour to secure the best possible regression equations, I carried out a number of stepwise multiple-regression analyses. In connection with a discussion of the complications of this method, I defined the significance requirements and emphasized the desirability of making cross-validations. The examination of the zero order correlations had already shown that several of the ability variables possessed high prediction values and the regression analyses performed also resulted in high multiple correlation coefficients and satisfactory *F*-values. The results also show that the variables in the sphere of psychomotor function are the best predictors. However, spatial-inductive and numerical abilities are also valuable as predictors and in the male group their values seem to be considerable. Finally, in this connection I was able to show that the combination of psychomotor variables with spatial-inductive and numerical variables generally resulted in a substantial increase in the accuracy of the predictions. These results were cross-validated in two respects. On the one hand, a parallel investigation was carried out in southern Sweden, in which, however, only values in the prognosis-group criterion could be obtained. On the other hand, the male group was divided into two halves and regression weights were calculated in each half. The calculated predicted criterion values were correlated with the observed values, and cross-validation coefficients were obtained varying between .48 and .57 in the parallel sample and between .57 and .74 in the two male sub-groups. It is true that the extent of shrinkage is considerable, as regards some of these coefficients, but all of them are significant at the 1 % level by adequate margins and, in my opinion, constitute final confirmations that it is possible to successfully predict working capacity in the mentally retarded from ability variables.

*Prediction of quantitative performances in eight tasks*

In my subsequent studies, I went on to analyse, by analogous methods, the possibilities of predicting the quantitative performances in the eight tasks selected. Here the situation was more favourable, in that it was possible to use as predictors not only measurements of ability but also observations of behaviour and combinations of these predictor groups. In the results section, I have already shown that the capacity of the measurements of ability to predict the quantitative work performances displays an obvious connection with the nature of the tasks, as it appears in their factor affiliations. Thus, among both men and women, the accuracy of prediction is greatest as regards the assembly of electrical panels and the making of wooden frames, which are the tasks which have the highest loadings in the first working-capacity factor. In these cases, the multiple correlation coefficients and *F*-values in the male group are practically as high as in the prediction of the prognosis-group and percentage-group criteria. If we make the same comparison in the female group, we find that the above-mentioned indices of prediction accuracy are even greater when the prediction refers to quantitative performances in the assembly of electrical panels and the making of wooden frames than when it refers to the prognosis-group and percentage-group criteria. The possibilities of predicting quantitative performances in other tasks from measurements of ability are by no means wanting but are not on the same high level and there is an apparent tendency for the predictability to decrease to the same extent as the tasks increase their loadings in the second working-capacity factor. We may obtain a numerical expression of this tendency by calculating the rank-difference correlations between the factor loadings and the multiple correlation coefficients. These rank-difference correlations are as high as .93 in the male group and .74 in the female group. If we consider only the most factor-pure tasks, however, this tendency to co-variation between factor affiliation and predictability appears strongest among the women.

The above-mentioned state of affairs—that it seems to be possible to predict from measurements of ability the quantitative performances in long-cycle and complex tasks (which also require the handling of tools) better than the corresponding performances in short-cycle and simple tasks—is an interesting result of these investigations and emphasizes that a further and more penetrating analysis of the dimensionality of work performances is urgently required, if we wish to gain a closer insight into the problems of prediction. That no attention has been paid to this state of affairs and that the mentally retarded have been practically exclusively observed "in the context of simple menial tasks which require little training, skill or attention, even for the retarded" (Gold 1973) may actually be one of the most important explanations of why researchers in previous prediction studies have comparatively seldom succeeded in achieving unambiguous results.

I found the possibilities of predicting quantitative performances from observations of behaviour surprisingly good, at any rate as regarded the male group. The differences between the tasks, as regards predictability, are smaller here, which would seem to be connected with the fact that subjective judgements are liable to halo effects. A certain relationship may be discerned between the predictability of the tasks and their factor affiliations, but this relationship is weaker than when the predictors consisted of measurements of ability. It is of particular interest to note that a certain shift has ensued in the ranking order between the tasks, as regards their predictability. The result of this is that observations of behaviour can predict, better than measurements of ability, quantitative performances in such tasks as the embroidering of cat motifs and the assembly of brackets, both of which were found to be assignable to the second working-capacity factor, while the state of affairs is the opposite as regards predicting performances in the assembly of electrical panels and the making of wooden frames, which had high loadings in the first working-capacity factor. This is important, as it means that the two predictor groups can effectively supplement each other, which was one of my initial hypotheses when I was planning these investigations. As I have already shown in the results section, considerable increases also took place in the accuracy of the predictions when I finally included measurements of ability and observations of behaviour in combination as independent variables in stepwise multiple-regression analyses.

In the female group, the results show that it is only passably possible to predict some of the quantitative performances from observations of behaviour. In some cases, the regression equations obtained are even quite unacceptable and in this respect it must be said that the difference between men and women is very obvious. However, when measurements of ability and observations of behaviour are combined as predictors, the effect is favourable even in the female group and results in certain increases as regards the proportions of explained criterion variance.

*Prediction of composite criterion variables*

I have previously maintained that the quantitative performances whose predictability I have just analysed can be regarded as satisfactory criteria of working capacity in the mentally retarded, since they are representative of the productive activity in the sheltered workshops designed for this group of individuals. Taken separately, however, they are, in a certain sense, comparatively specific as criteria. Each of them includes only the demands on the workers which derive from the nature of the individual task. It was this fact which justified my also calculating, in order to obtain a criterion variable which would represent, in a more comprehensive fashion, the working capacity of the mentally retarded, via transformations into $z$ values the averages of the quantitative performances in the

eight tasks and regarding these measurements as a main criterion. When I also developed additional composite criterion variables, I did so in order to be able to distinguish three main aspects of the working capacity—the quantitative aspect, the qualitative aspect and an aspect referring to adjustment to work—and to analyse the possibilities of predicting each of these aspects. The results of these analyses, which will be found in the later part of Chapter V, are quite compatible with what we have found earlier, while in certain respects they give us greater insight into the problems of prediction.

The fact that the criterion variable entitled "Average quantitative performance" is more heterogeneous or multi-dimensional than the quantitative performances in the respective tasks does not seem to have had any inhibiting effect worth mentioning on the prediction possibilities. Thus, this criterion variable can be successfully predicted from measurements of ability. This proves to be the case among both men and women, and the only observable difference between the sexes is the women's lower *F*-values. On the other hand, when the predictors consist of observations of behaviour, we find an obvious difference between men and women—just as was the case in the prediction of quantitative performances in different tasks. Thus, in the male group, the regression equation obtained is much more satisfactory than in the female group. Conditions fully analogous to those just reported apply in the prediction of criterion variable 3:8, " Average rating of quantitative performance", which reminds us that this variable reflects to a large extent the objective results of measurement which are the main criterion.

If we turn our attention from these two criterion variables, which measure especially quantitative performances, to the three variables which instead measure the work performances from the qualitative viewpoint, we shall have reason to observe once more what we have already noted in the results section, viz. that in their case the predictability seems to be somewhat greater. This applies independently of which predictor group we use and the tendency is clearly marked among both men and women. Just as, judging from the previously discussed results, it is easier to predict quantitative performances in long-cycle and qualitatively more demanding tasks, it seems to be possible to predict the qualitative aspects of the performances with somewhat greater success than the quantitative aspects. It appears to be likely that we can explain both these circumstances, at least in part, as a consequence of variations in the homogeneity or dimensionality of the criterion variables. In the factor analysis on p. 95 of the correlations between the composite criterion variables, the especially quantitatively determined performances were found to have higher loadings in the adjustment-to-work factor than the especially qualitatively determined performances. Similarly, we were earlier able to discern in the quantitative performances in such tasks as cutting threads in nuts and assembling brackets a certain

tendency to higher correlations with the adjustment criteria. Thus, in both these cases, there is evidence that the factor structure is more complex in the especially quantitatively determined composite criterion variables, as well as in the quantitative performances, which are to be assigned to the second working-capacity factor, and I hold that it is probable that this state of affairs contributed to the lower degree of predictability in these variables. In spite of the fact that I supplemented the measurements of ability with observations of behaviour, I was not, to all appearances, able to design in these cases a predictor structure which would in all respects correspond to the complex criterion structure.

Finally, as regards the composite adjustment criteria, I have shown in the foregoing pages that it is scarcely possible to predict these criteria from measurements of ability. On the other hand, in this connection observations of behaviour may serve as satisfactory predictors among both men and women. In certain cases, a combination of measurements of ability and observations of behaviour seems to lead to a slight increase in the accuracy of the predictions. The results emphasize the importance of further improvements in the methods of observation, since this seems to offer the only practicable solution when it is desired to predict adjustment to work in the mentally retarded.

*The trustworthiness of the results*

The validity of the procedures applied is indicated primarily by the fact that in all the regression analyses I could calculate indices of prediction accuracy, which were found to be of an almost satisfactory or—in the majority of cases—of a highly satisfactory magnitude. However, the cross-validations should be assigned even greater weight in this context. In the first stage these referred to the procedure which was applied in the prediction of the prognosis-group and percentage-group criteria and resulted in cross-validation coefficients which constitute the most positive proof so far that it is possible to successfully predict working capacity in the mentally retarded on the basis of measurements of ability. However, I was anxious to extend the cross-validation investigations to refer also to the somewhat more complicated procedure used in the prediction of quantitative performances in the tasks and composite criterion measures. As I have shown in the foregoing pages, these investigations also resulted in cross-validation coefficients which merit the description "unusually high".

Thus, my investigations end in the conclusion that it is possible to successfully predict work performances and adjustment to work in mentally retarded adults from combinations of measurements of ability and observations of behaviour, regardless of the criterion variables used (prognosis-group and percentage-group criteria, quantitative performances in individual tasks or composite performance and adjustment criteria). These positive results may certainly appear somewhat surprising in the light of the research tradition in this field described in the in-

troduction, a tradition which contains so many disappointed hopes and contradictory conclusions. On this account, there is also particular reason to ask whether false correlations or other artefacts may have contributed to increase the accuracy indices in the stepwise multiple-regression analyses or may have helped to embellish in other ways the values obtained.

When the first results were available, I asked myself whether the spread in the experimental group might have been so great as to make the correlations obtained unrealistically high. However, it is scarcely possible to maintain that this was the case, if one bears in mind that the spread of the IQ was limited to 11.04 units in the male group and 8.86 units in the female group. The corresponding figures in the somewhat reduced groups which took part in the testing programme in the sheltered workshop were 10.38 and 8.51 respectively. I have previously shown that these values are equal to or lower than the standard deviations in whole classes of pupils who finish their compulsory schooling in a special school. It is evident that they are much smaller than the corresponding values in groups of normal pupils. However, one may return to the argument with the refuted objection in a somewhat modified form and assert that the investigations took place within such continua of the dimensions of behaviour that differences between individuals appear to a more marked extent than they otherwise would. This viewpoint may be justified and is in any case worthy of further attention. In this connection, I may recall what I pointed out earlier (cf. p. 40), that, in certain psychomotor tests, comparatively large gaps appear towards the negative end of the distribution and that, in some cases, tendencies to bimodal distributions occur. However, it is unlikely that such circumstances would have had any considerable influence on the reported results.

Thus, I consider that I may assert that the positive results must be primarily considered as manifestations of the favourable research conditions and the associated facilities fulfilling the basic methodological requirements which are of topical interest in all investigations of predictive validity. As regarded the selection of predictors, I was able to benefit to a large extent from previous investigations, which do not by any means contain only inconsistencies and failures but also contain quite a number of lucid analyses and promising results, as is clear from the introduction. In the ability measurements, I had the assistance of psychologists who were not only trained in testing methods but also took part in the practical work of habilitation and acquired in the course of it a good knowledge of the special characteristics of those who took part in the experiments. The actual tests were carried out under satisfactory conditions and also resulted in values which were shown to have satisfactory reliability. The observations of behaviour in working situations were made by supervisors who were interested in the task and had been well instructed. These observations also resulted in predictors which fulfil reasonable requirements as regards reliability.

Finally, the fact that I was able to use the collective facilities of a well-managed workshop for two whole years, in order to analyse the structure of the criterion variable and to collect reliable measurements of different aspects of this variable would seem to have been of even greater importance for the positive results of the investigation. There is here reason to assert once more that neither observational nor testing programmes could be carried out without the continual and stimulating co-operation with the workshop manager, the research technician and the supervisors. My experience in this respect seems to be in close agreement with Gold's comments on the relationship between the researchers and practitioners (Gold 1973).

## 2. Unanswered Questions and Needs of Further Research

While I also feel justified in asserting that the investigations described in the preceding pages convincingly demonstrate the possibilities of successfully predicting work performances and work adjustment in the mentally retarded, and that this main purpose has thereby been achieved, a number of questions nevertheless remain unanswered, which at the same time constitute important subjects for future research. Some of these will be touched on in the following pages.

*Identification of valid predictors and analysis of predictor structures*
One of the essential subsidiary results of these investigations is the identification of valid predictors in the form of ability measurements and observational data. Yet it is obvious that further progress may be made in this field by doing more work on test construction. One of the first aims of such work should be to modify and complement certain of the methods here applied, in order that they may be better adjusted to the mentally retarded. The measurements of intra-subject variability, which were included in the regression analyses reported here, have earned their place and are worthy of further attention. At the beginning of the investigations, certain attempts were made to use measures of the improvement rate as predictors, as had been suggested by several other researchers. However, the results were hardly favourable and the attempts were abandoned until further notice. Nevertheless, in my opinion there is some justification for following up the question of whether such measures can make any specific contribution to the prediction of the criterion variable.

The physical capabilities of the experimental subjects in this investigation have been studied by Nordgren (1970). This author has been able to show that there are substantial correlations between muscle-strength measurements and preliminary versions of the prognosis-group and percentage-group criteria (cf.

Nordgren and Bäckström 1971). It should be of great interest to combine such clinical-physiological data as predictors with ability measurements and observations of behaviour. Finally, on the basis of the analyses of the predictor structure, already performed, there is reason to investigate whether combinations of factor scores can be more effective predictors than combinations of separate test scores.

*The dimensionality of the criterion variable*

The task of further clarifying the dimensionality of the criterion variable is even more urgent. The efforts to this end reported in this book are related to current trends in industrial psychology, according to which one ought to validate selection methods against job performance rather than against supervisory ratings, while realising also that such a criterion is multi-dimensional in nature (Ronan and Prien 1966, Bray and Moses 1972). Accordingly, I was able to show not only that the eight tasks, which together made up a test programme, can, more or less definitely, be referred to either of two factors of working capacity, but also—which is the most interesting thing from the point of view of prediction—that there is an obvious relationship between the factor affiliation of the tasks and their predictability. Moreover, several examples have emerged of the more or less self-evident fact that the effective predictor structure changes character, when the criterion structure is changed. The researcher with the necessary resources at his disposal may well be attracted to the task of carrying out, on a sufficiently large scale, "factor-structure-seeking studies", which are intended to finally lay bare the factor structure inherent in the working capacity, and to relate the results thereby obtained to prediction possibilities and predictor structures. It seems by no means impossible that the completion of such a research programme could be of value, not only as a basis for the individual guidance of the mentally retarded but also as a model for similar investigations of normal workers. In his review of research, Windle developed the defensible view that "investigation of subnormals would yield more useful information about normals, because certain complicating features of normals are more or less controlled in the subnormal" (Windle 1962).

*Longitudinal studies and long-range prediction*

There is justification for asserting that the investigations reported in this book refer throughout to predictive validity, since, as I have already mentioned, the predictor data were invariably collected before the criterion values. However, the time-lag between these two occasions was comparatively short and could not be made the same for all experimental subjects, as was desirable. As regards this question, I may sum up by saying that the ability measurements were made 1–2 years before the criterion measurements. On the other hand, the period between

the behaviour observations and commencement of criterion measurements was no longer than 2–4 weeks. Insofar as the finding of a basis for long-range prediction is regarded as a matter of urgency, these conditions constitute obvious weaknesses. I realized this and I have already made preparations for new studies, which are intended to be longitudinal in character and are designed to study methods of using the results obtained in connection with the individual guidance of mentally retarded youngsters who have completed their compulsory schooling (Palmér 1970). Such studies are urgent, as the road which leads from regression equations to placement decisions regarding individual young people is a long one, paved with many unanswered questions (cf. Cronbach 1971). Moreover, my analyses of the structure of the criterion variable will in the process be extended to cover competitive work situations, which are obvious goals for the less seriously retarded. I am well aware that the whole problem complex thus becomes considerably more complicated. There are many indications that in such a longitudinal study there is reason to allow the eight tasks to change roles and serve as predictors instead of as criteria.

*Sex differences*

The question of the similarities and differences between men and women has made its presence felt in various connections in the investigations reported here. Psychologists are tempted to turn a blind eye to this question, since it is complicated and rarely permits of a definitive answer. The investigator who studies the mentally retarded is especially inclined to regard men and women as one and the same group, because he encounters particularly great difficulties in his laudable efforts to observe a sufficiently large number of experimental subjects. However, there are good reasons, from both the theoretical and the practical point of view, to note sex differences in this group. Some observations of Wessman and Gustin (1964) support this. As I mentioned above, these authors studied four grades (born between 1935 and 1939) of the total number of pupils attending special schools in Sweden and found that the distribution was 60 % men and 40 % women. Moreover, they were able to establish that the average IQ was consistently lower in women than in men. On admission to the special school, the difference was 3.3 units, and by discharge it had increased to 5.5 units. Such a difference was found also in the sub-groups of men and women, who were assessed, after varying periods of probation at the special school, as lacking in the capacity to benefit by the instruction and were therefore transferred to an institution for residential care. Thus, the indications for placement in special school or institution were considered to be on a higher intellectual level among boys than among girls.

The causal relationships underlying the circumstances described here cannot be regarded as having been elucidated. Is it true that women are more easily

accepted as worthy members of the family and society, while greater demands are made on men? This hypothesis is supported by certain other observations made by Wessman and Gustin, namely, that the women—when they reach maturity—exhibit a higher frequency of marriage and pregnancy, while the men are better able to earn a living, a circumstance expressed in their greater tendency to join trade unions and receive contractual wages. However, other hypotheses may also be justified. The regulations provide that children may be assigned to special schools, if they are considered to lack the capacity to benefit by education in the primary school or its special classes. Is it simply that girls are able to benefit by the instruction provided by the primary school on a lower intellectual level than that applying to boys? Wessman and Gustin in fact also report observations which support such a hypothesis. They find that boys and girls do not reveal any differences to speak of as regards knowledge of reading, writing and arithmetic, despite the above-mentioned differences as regards intelligence.

In connection with the studies reported here I have asked whether they can, in some measure, throw light upon the questions that arise from the observations of Wessman and Gustin. At first I noticed that the correlations between the rating criteria and the objectively determined work performances were lower in women than in men. This was also the case when comparing the correlations between the predictor variables in the form of observations of behaviour and the objectively determined work performances. A possible interpretation was to regard these circumstances as an indication that the judgement processes themselves are more complicated when applied to women than when applied to men. For example, it could be that these processes were influenced by different attitudes toward the behaviour of men and women in their working life, or, on the other hand, that female behaviour was influenced more by situational factors and that judgements of that behaviour therefore displayed a lower degree of reliability.

However, the correlations between the scores of the ability variables and the quantitatively determined work performances were also lower in women than in men. Furthermore, later on in the study, the likelihood of correctly predicting the three adjustment criteria from the observations of behaviour proved to be better in the female group than in the male group. The latter circumstance demonstrates that it is quite possible to make reliable observations of female behaviour. However, what seems to be associated with obvious difficulties is to accurately predict factual work performances from observations of behaviour.

If we wish to trace the causes of these difficulties, we have to bear in mind that factual work performances are normally determined by an interaction between individual, motivational and situational factors. Is it perhaps the case that this interaction process does not take place in the same manner in men and women and that the structure of behaviour itself, thus reveals sexual differences?

Some observations in the studies reported here support such a hypothesis. In

scrutinizing the correlations between the observation variables and the eight composite criterion measures I found, thus, that the motivation variables in the male group tended to occupy an intermediate position on a continuum extending from a markedly emotional reaction to a primarily cognitive or psychomotor performance while those in the female group seemed to lie nearer the emotionally loaded pole in the assumed continuum. I also noted a corresponding tendency in the intercorrelations between the sub-scales in the general observation schedule of the check-list type.

I have already pointed out that it is not possible to reach any definite conclusions as regards the differences between men and women in a study like this one. Nor can the observations noted here reveal in any definitive sense the underlying causes of the observations reported by Wessman and Gustin. However, as the above discussion shows,they can provide the basis for further inquiries into these questions, which it is essential to clarify as far as possible, since they illustrate important factors in the habilitation process and its consequences for the individual, For similar reasons, I was unwilling to neglect making consistent comparisons between men and women in other respects too. In the following pages I shall mention by way of summary the observations made and shall first consider the prediction variables.

In measurements of the ability variables men and women displayed for the most part the same average standard of performance. Nevertheless, it was confirmed that the men had better results in certain psychomotor and spatial tests, while the women, on the other hand, tended to be more successful in numerical tasks. Observations showing the same trend have been made by other researchers. In a number of cases, the intercorrelations proved to be higher among men than among women, which, however, could, at least to some extent, be interpreted as a consequence of the lower variability in the female group. Later factor analyses also showed marked similarities between the factor structure among men and women, insofar as a numerical, a spatial-inductive, and a psychomotor factor with the same content could be distinguished in both groups. These results confirm what Kebbon reported previously. As regards a second psychomotor factor and a reaction-time factor, the agreements between men and women were not as complete, which indicates that certain structural differences will probably emerge if the whole of the extensive sphere of psychomotor function is analysed.

So far as performance in the ability variables is concerned, the similarities between men and women are very striking. The differences which could be noted are, of course, of interest and worthy of further study but can rather be designated as marginal. Nor are the differences great in the observation variables, which refer to working capacity chiefly from the qualitative point of view. Notwithstanding that they do not reach statistical significance, they cannot be

neglected, as the men consistently show somewhat higher values than the women. Finally, the differences which emerge in those observation variables which measure work behaviour from the emotional and motivational viewpoints are the most striking. Here the men show consistently higher values and in most cases the differences reach a statistically significant level. In view of the facts now established, I can sum up by saying that the differences between men and women apparent in the prediction variables tend to become more marked, insofar as the comparisons concern more modifiable qualities.

Such a tendency also seems to emerge from a consideration of the sex differences in the criterion variables. With one exception, the men show higher values throughout than the women. The differences with regard to qualitative work performance, however, are comparatively slight and do not reach a statistically significant level. The differences are more obvious in the case of quantitative work performances. The account of the eight tasks shows that the men performed significantly better in the making of wooden frames, the turning of holders and the assembly of electrical panels. In the cutting of threads in nuts, the assembly of brackets, the pressing of eyelets and the packing of blocks, the differences showed the same direction, although without being significant. Embroidering cat motifs was the characteristic exception, in that the women here achieved a significantly better performance than the men. The sex difference in the criterion variable "Average quantitative performance" may be regarded as summing up the difference between men and women, as regards quantitative performance. Here, of course, the men are more or less clearly superior, although the difference does not attain the level of statistical significance. Its $t$-value is 1.75. Finally, as regards the three adjustment variables, significant differences in them appear consistently, indicating that the women are characterised by a lower degree of adjustment to work than the men.

Thus, a similar pattern recurs in both predictors and criteria. In the ability variables, as in variables measuring work performance primarily from the viewpoint of quality, the sex differences are slight. In variables which refer to quantitative work performance, the magnitude of the differences increases. Finally, they are most marked in variables measuring emotional and motivational aspects of behaviour.

This pattern means, as I have just stated, that the differences between men and women tend to be the more marked the more the comparisons refer to modifiable qualities. This circumstance is in itself an indication that the origin of the differences should, to a considerable extent, be sought in differences as regards earlier training and previous experience. It may be of importance to make this clear, since it means that the differences described would probably be considerably less if the women were to receive the same training and the same stimulation as the men during their upbringing. On the other hand, this must not

lead us to turn a blind eye to, or neglect, the differences which do in fact exist, and which in these investigations are most clearly apparent in the adjustment variables.

*Prediction research and the effects of training*

The view has been put forward from several quarters that we should not waste our time studying evaluation and prediction among the mentally retarded, but should instead attack problems related to training and the effect of training. Thus, representatives of the parents' associations have on occasions devised such slogans as "Don't test them but train them". Even some psychologists have concurred and taken advantage of the opportunity to make known their dissatisfaction with intelligence tests and measurements of behaviour in general. When Gold (1973) recommends that prediction investigations of the mentally retarded should be suspended, at least for the time being, he does so on the basis of more factual *points de départ*. He is pessimistic as regards the possibilities of developing predictors and using them "to predict success on a particular job or task"; he also seems to be in favour of awaiting social developments which will provide the mentally retarded with more differentiated opportunities and tasks. For my part, I am inclined to say that the results reported here justify a more positive attitude to prediction studies even at this stage. Above all, there is reason to claim that investigations of this kind may provide incentives to social developments and thereby in their turn help to bring about better working conditions for the mentally retarded, as well as better conditions for further research.

However, it is more urgent to underline the vital importance of maintaining a connection, both in theory and in practice, between the aspects of prediction and training than to defend the justification of limited prediction studies. The expression *diagnostic teaching* has been used in the sphere of special education to designate "the tendency to integrate assessment, training and control of consequent effects into one unit" (Österling 1967). Such an attempt at integration is equally essential in connection with the habilitation of the mentally retarded. The assessment procedures are here intended to form a basis for a treatment programme, and their evaluation then becomes a *point de départ* for further and modified treatment in a succeeding phase.

However, successful prediction investigations in themselves presuppose that we shall, in more than one respect, devote attention to the conditions and effects of training. This has been exemplified several times in the studies reported in this book and my final task in this context will be to summarize and comment on these examples.

In connection with the measurements of the quantitative performances in the eight tasks, the question arose of how long a period of preparatory instruction and training should be afforded the experimental subjects. As I have reported in

detail in the preceding pages, the answer was determined, on the one hand, by the desire to offer a sufficiently long training period and, on the other, by such practical problems as access to workplaces, financial resources, etc. Partly by reason of the impossibility of extending the work period to cover more than four days on each task, I judged it advisable to study the development of behaviour in detail throughout this period and, *inter alia,* to record it in the form of a diagram of the development of performance. The preceding pages also include a general analysis of this development—from the point of view of both quality and quantity. Improvements in quality could only be noted in some of the tasks and were evidently most closely related to the degree of difficulty which characterized them from the beginning. In terms of quantity, the effects of training emerged clearly in all the tasks, except the embroidering of cat motifs. Their extent was greatest at the beginning of the working week. If we disregard the embroidering of cat motifs—which seems to differ on several points from the other tasks—the quantitative contributions of training also co-vary with the degree of difficulty, inasmuch as their extent becomes greater in proportion to the increases in the degree of difficulty. The calculated rank-difference correlation between the variables amounts to 0.71 among the men and to 0.75 among the women. Side by side with the training effects, another systematic trend could be demonstrated in certain of the tasks, which meant that the performances were lower in the afternoon than in the morning sessions. It is possible that a more far-reaching analysis of the qualitative and quantitative developments would yield results of considerable interest—not least as regards the occurrence of individual differences. However, such an analysis involves complicated statistical problems and I have not yet been able to carry it out.

One of my particular reasons for studying the development of behaviour during the training periods was to obtain a basis on which to assess the important question of whether there was any risk that the performance measurements used would undergo any substantial changes in connection with further training and consequently be unsuitable, in their present form, to serve as good criteria. This question is discussed in a separate appendix, where it is answered, with certain reservations, in the negative. Instead, I have given reasons for regarding these measurements of performance as sufficiently stable and realistic expressions of the working capacity of the experimental subjects. In addition, I have put forward the hypothesis that in the long run variations in the work performance are determined by factors of motivation rather than of training.

Further light is shed on the question of the long-term development of performance by the control study reported in Appendix D, concerning the extent of possible transfer effects in connection with the measurements of the quantitative performance in the tasks. It was there stated that signs of positive transfer effects only appeared in the making of wooden frames. At the same time, this investiga-

tion offered some support for the hypothesis that negative effects, resulting from incipient fatigue or waning motivation, arose instead in certain of the other tasks. However, none of the differences which agree with this hypothesis attain a significant order of magnitude, so that there are no grounds for definite conclusions. Nevertheless the observations made in these investigations make it important that, in future long-term studies of the development of performance among the mentally retarded, as much attention should be devoted to motivation factors as to training factors.

# Appendices

# Appendix A

*Appendix A:1.* Correlation matrix for 27 prediction variables. Men, $n = 47$.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Crawford screws | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. Crawford pins | .69 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. O'Conner test | .70 | .74 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. Mark-making | .43 | .54 | .51 | | | | | | | | | | | | | | | | | | | | | | | | |
| 5. Wire labyrinth | −.57 | −.62 | −.55 | −.61 | | | | | | | | | | | | | | | | | | | | | | | |
| 6. Threading | −.60 | −.59 | −.49 | −.62 | .74 | | | | | | | | | | | | | | | | | | | | | | |
| 7. Plates | .52 | .66 | .60 | .60 | −.55 | −.50 | | | | | | | | | | | | | | | | | | | | | |
| 8. MTTA | −.54 | −.70 | −.66 | −.73 | .74 | .61 | −.64 | | | | | | | | | | | | | | | | | | | | |
| 9. Tapping | −.43 | −.48 | −.50 | −.60 | .80 | .65 | −.54 | .57 | | | | | | | | | | | | | | | | | | | |
| 10. Copying | .57 | .68 | .57 | .64 | −.56 | .42 | .62 | −.73 | −.45 | | | | | | | | | | | | | | | | | | |
| 11. Reaction time | −.30 | −.51 | −.34 | −.38 | .29 | .24 | −.42 | .37 | .23 | −.31 | | | | | | | | | | | | | | | | | |
| 12. Memory for design | .24 | .53 | .43 | .40 | −.24 | −.25 | .51 | −.33 | −.27 | .45 | −.20 | | | | | | | | | | | | | | | | |
| 13. Surface developm. | .34 | .61 | .48 | .34 | −.28 | −.23 | .45 | −.47 | −.17 | .65 | −.29 | .52 | | | | | | | | | | | | | | | |
| 14. Block design | .48 | .69 | .59 | .42 | −.28 | −.33 | .69 | −.50 | −.30 | .67 | −.40 | .59 | .69 | | | | | | | | | | | | | | |
| 15. Raven's Matrices | .27 | .43 | .46 | .46 | −.25 | −.13 | .67 | −.46 | −.25 | .62 | −.32 | .54 | .63 | .77 | | | | | | | | | | | | | |
| 16. The Leiter Scale | .51 | .61 | .68 | .65 | −.48 | −.47 | .66 | −.62 | −.50 | .68 | −.40 | .55 | .55 | .74 | .69 | | | | | | | | | | | | |
| 17. Figure classification | .35 | .51 | .47 | .51 | −.37 | −.35 | .73 | −.52 | −.42 | .61 | −.39 | .52 | .62 | .78 | .77 | .71 | | | | | | | | | | | |
| 18. Concepts | .48 | .53 | .65 | .60 | −.39 | −.49 | .49 | −.68 | −.43 | .65 | −.32 | .31 | .48 | .61 | .49 | .72 | .58 | | | | | | | | | | |
| 19. Addition, UB | .18 | .31 | .35 | .52 | −.31 | −.31 | .54 | .54 | −.40 | .48 | −.21 | .38 | .24 | .55 | .56 | .60 | .60 | .54 | | | | | | | | | |
| 20. Addition, W | .21 | .45 | .50 | .57 | −.40 | −.39 | .61 | −.56 | −.47 | .62 | −.36 | .43 | .45 | .60 | .64 | .71 | .68 | .66 | .83 | | | | | | | | |
| 21. Subtraction, W | .06 | .38 | .36 | .50 | −.31 | −.29 | .51 | −.43 | −.38 | .57 | −.35 | .44 | .44 | .58 | .58 | .67 | .66 | .54 | .82 | .88 | | | | | | | |
| 22. Highest number | .36 | .47 | .46 | .57 | −.40 | −.36 | .70 | −.49 | −.42 | .51 | −.40 | .40 | .40 | .70 | .70 | .70 | .80 | .60 | .81 | .77 | .71 | | | | | | |
| 23. Identical forms | .31 | .50 | .49 | .58 | −.44 | −.38 | .76 | −.62 | −.52 | .68 | −.27 | .42 | .47 | .69 | .65 | .66 | .75 | .59 | .61 | .75 | .63 | .69 | | | | | |
| 24. Vocabulary | .32 | .43 | .44 | .54 | −.36 | −.36 | .44 | −.54 | −.46 | .66 | −.13 | .42 | .55 | .53 | .50 | .65 | .49 | .73 | .40 | .49 | .42 | .45 | .52 | | | | |
| 25. Reaction time, S.D. | −.34 | −.45 | −.35 | −.30 | .24 | .17 | −.44 | .31 | .21 | −.21 | .82 | −.15 | −.22 | −.39 | −.38 | −.28 | −.41 | −.25 | −.24 | −.36 | −.33 | −.41 | −.28 | −.03 | | | |
| 26. Wire labyrinth, S.D. | −.48 | −.62 | −.55 | −.54 | .89 | .55 | −.54 | .79 | .72 | −.60 | .42 | −.25 | −.33 | −.32 | −.32 | −.53 | −.39 | −.44 | −.26 | −.41 | −.32 | −.38 | −.42 | −.38 | .33 | | |
| 27. Tapping, S.D. | −.28 | −.40 | −.45 | .44 | .71 | .34 | −.38 | .59 | .72 | −.39 | .15 | −.18 | −.01 | .11 | .09 | −.35 | −.18 | .23 | −.22 | −.28 | .17 | .20 | −.29 | −.24 | .12 | .71 | |

*Appendix A:2.* Correlation matrix for 27 prediction variables. Women, $n = 33$.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Crawford screws | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. Crawford pins | .44 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. O'Connor test | .68 | .49 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. Mark-making | .34 | .18 | .48 | | | | | | | | | | | | | | | | | | | | | | | | |
| 5. Wire labyrinth | −.53 | −.34 | −.54 | −.28 | | | | | | | | | | | | | | | | | | | | | | | |
| 6. Threading | −.52 | .40 | −.67 | .27 | .58 | | | | | | | | | | | | | | | | | | | | | | |
| 7. Plates | .70 | .64 | .52 | .38 | −.46 | −.52 | | | | | | | | | | | | | | | | | | | | | |
| 8. MTTA | −.49 | −.52 | −.39 | −.38 | .46 | .48 | −.56 | | | | | | | | | | | | | | | | | | | | |
| 9. Tapping | −.33 | −.39 | −.30 | −.42 | .28 | .39 | −.62 | .60 | | | | | | | | | | | | | | | | | | | |
| 10. Copying | .56 | .58 | .38 | .30 | .43 | .43 | .53 | −.75 | .52 | | | | | | | | | | | | | | | | | | |
| 11. Reaction time | −.41 | −.28 | −.19 | −.43 | .34 | .31 | −.63 | .44 | .61 | −.46 | | | | | | | | | | | | | | | | | |
| 12. Memory for design | .18 | .28 | .14 | .18 | −.34 | −.25 | .29 | −.29 | −.42 | .50 | −.22 | | | | | | | | | | | | | | | | |
| 13. Surface developm. | .35 | .43 | .30 | .34 | −.18 | −.14 | .24 | −.41 | −.24 | .59 | −.38 | .15 | | | | | | | | | | | | | | | |
| 14. Block design | .19 | .25 | .14 | .15 | −.35 | −.29 | .36 | −.35 | −.40 | .58 | −.30 | .76 | .36 | | | | | | | | | | | | | | |
| 15. Raven's Matrices | .11 | .25 | −.01 | .01 | −.21 | .02 | .18 | −.16 | −.28 | .59 | −.36 | .50 | .47 | .65 | | | | | | | | | | | | | |
| 16. The Leiter Scale | .22 | .41 | .11 | .15 | −.16 | −.10 | .35 | −.29 | −.50 | .59 | −.35 | .40 | .46 | .54 | .55 | | | | | | | | | | | | |
| 17. Figure classification | .18 | .40 | .15 | .24 | −.37 | −.16 | .46 | −.25 | −.40 | .54 | −.45 | .49 | .15 | .55 | .55 | .54 | | | | | | | | | | | |
| 18. Concepts | .20 | .29 | .19 | .21 | −.09 | −.02 | .31 | −.19 | −.58 | .41 | −.33 | .25 | .33 | .39 | .42 | .69 | .53 | | | | | | | | | | |
| 19. Addition, UB | .30 | .16 | .32 | .27 | −.21 | −.13 | .26 | −.24 | −.41 | .30 | −.29 | .24 | .21 | .24 | .37 | .46 | .26 | .36 | | | | | | | | | |
| 20. Addition, W | .24 | .23 | .21 | .35 | −.14 | −.07 | .36 | −.33 | −.56 | .52 | −.42 | .40 | .47 | .57 | .54 | .65 | .52 | .65 | .73 | | | | | | | | |
| 21. Subtraction, W | .05 | .12 | .12 | .32 | −.08 | .10 | .18 | −.17 | −.35 | .37 | −.25 | .38 | .44 | .49 | .65 | .54 | .45 | .44 | .72 | .85 | | | | | | | |
| 22. Highest number | .10 | .15 | .12 | .20 | .13 | .06 | .18 | −.28 | −.34 | .48 | −.23 | .32 | .26 | .43 | .65 | .50 | .52 | .48 | .56 | .73 | .81 | | | | | | |
| 23. Identical forms | .32 | .33 | .36 | .28 | −.13 | −.22 | .31 | −.25 | −.38 | .54 | −.19 | .15 | .37 | .35 | .49 | .37 | .45 | .42 | .42 | .57 | .48 | .51 | | | | | |
| 24. Vocabulary | .16 | .26 | .28 | .36 | .03 | −.19 | .27 | −.31 | −.55 | .55 | −.46 | .33 | .53 | .49 | .44 | .59 | .44 | .73 | .36 | .71 | .54 | .48 | .50 | | | | |
| 25. Reaction time, S.D. | −.21 | .37 | −.11 | −.15 | .15 | .16 | −.47 | .42 | .57 | −.37 | .69 | −.21 | −.28 | −.35 | .39 | −.25 | −.43 | −.38 | .19 | −.42 | −.25 | −.26 | −.31 | −.41 | | | |
| 26. Wire labyrinth, S.D. | −.38 | −.30 | −.44 | −.17 | .79 | .32 | −.32 | .23 | .26 | .35 | .34 | −.34 | −.19 | −.34 | −.29 | −.14 | −.39 | −.14 | .16 | −.13 | .04 | −.04 | −.22 | −.13 | .24 | | |
| 27. Tapping, S.D. | −.23 | −.43 | −.33 | −.22 | .28 | .28 | −.49 | .54 | .64 | −.41 | .47 | −.44 | −.20 | .47 | −.30 | −.22 | −.36 | −.31 | −.25 | −.35 | −.17 | −.16 | −.31 | −.28 | .74 | .37 | |

# Appendix B

*Appendix B:1.* Sub-scales in the observation of working behaviour. $n = 66$ (40 men and 26 women)

I. Emotional Stability

| Check-list no. | Serial no. | Wording of item | $r_{pbi}$ | $p$ value |
|---|---|---|---|---|
| 15 | 1 | Is emotionally calm and steady | 0.83 | 0.68 |
| 128 | 2 | Can control his or her feelings to a remarkably small extent | 0.78 | 0.27 |
| 69 | 3 | Is quite unbalanced on some days and then works very little | 0.75 | 0.30 |
| 130 | 4 | Seldom complains about anything | 0.74 | 0.59 |
| 93 | 5 | Is unpredictable and changes his or her attitude from one moment to another | 0.73 | 0.24 |
| 79 | 6 | Is remarkably sensitive to criticism from others | 0.69 | 0.48 |
| 7 | 7 | Is touchy and irritable | 0.68 | 0.38 |
| 31 | 8 | Is sometimes distressed and cries at setbacks | 0.66 | 0.32 |
| 26 | 9 | Endures setbacks better than most | 0.61 | 0.36 |
| 133 | 10 | Seems able to remain calm and assured in practically all situations | 0.61 | 0.36 |
| 74 | 11 | Seems to be sensitive to noise | 0.59 | 0.11 |
| 68 | 12 | Often complains of headache or stomach ache | 0.58 | 0.17 |
| 6 | 13 | Cries easily | 0.58 | 0.14 |
| 132 | 14 | Easily becomes worried and anxious when something unexpected happens | 0.57 | 0.45 |
| 5 | 15 | Displays an even and calm temper in the work situation | 0.56 | 0.77 |
| 153 | 16 | Is very successful at tasks in which he can put his back into the work and exert his physical strength | 0.54 | 0.45 |
| 17 | 17 | Is restless in the work situation | 0.52 | 0.29 |
| 142 | 18 | Can work concentratedly even in unfamiliar work situations | 0.51 | 0.58 |
| 60 | 19 | Becomes nervous and uneasy in study visits with many participants | 0.51 | 0.15 |
| 49 | 20 | Needs special encouragement in order to get him or her to start work in the morning | 0.49 | 0.15 |
| 109 | 21 | Often expresses anxiety and alarm about the future | 0.47 | 0.15 |
| 84 | 22 | Is sometimes self-absorbed and apathetic | 0.44 | 0.20 |
| 127 | 23 | Tries to cope with a difficult task himself or herself, rather than ask for help at once | 0.34 | 0.48 |
| 37 | 24 | Seems almost always unhappy and depressed | 0.31 | 0.09 |

II. Ability to Co-operate, with Absence of Tendencies to be Aggresive and to Domineer.

| Check-list no. | Serial no. | Wording of item | $r_{pbi}$ | $p$ value |
|---|---|---|---|---|
| 137 | 1 | Is often displeased and dissatisfied | 0.77 | 0.32 |
| 115 | 2 | Is sometimes noisy and boisterous | 0.77 | 0.35 |
| 50 | 3 | Seldom quarrels with fellow-workers | 0.73 | 0.59 |
| 51 | 4 | Readily raises objections | 0.73 | 0.29 |
| 97 | 5 | Practically never falls out with his or her fellow-workers | 0.70 | 0.61 |
| 92 | 6 | Creates uneasiness in those around him or her | 0.70 | 0.17 |
| 73 | 7 | Complains that the working day is long | 0.67 | 0.24 |
| 83 | 8 | Likes to domineer over his or her fellow-workers | 0.66 | 0.21 |
| 129 | 9 | Comes to blows with fellow-workers on very rare occasions | 0.65 | 0.17 |
| 72 | 10 | Is sometimes aggressive and threatening to fellow-workers | 0.65 | 0.17 |
| 135 | 11 | Helps to create a good atmosphere in the workplace | 0.65 | 0.59 |
| 113 | 12 | Likes to argue and discuss things | 0.64 | 0.47 |
| 32 | 13 | Becomes angry and shows bad temper at setbacks | 0.61 | 0.44 |
| 38 | 14 | Is self-assured and over-estimates own ability | 0.60 | 0.21 |
| 101 | 15 | Is anxious to fulfil all his or her responsibilities at the work-place | 0.59 | 0.73 |
| 118 | 16 | Frequently talks a lot | 0.58 | 0.40 |
| 8 | 17 | Co-operates well with fellow-workers | 0.56 | 0.79 |
| 70 | 18 | Sometimes falls into fits of rage and throws things | 0.51 | 0.12 |
| 33 | 19 | Often shirks common tasks | 0.48 | 0.11 |
| 48 | 20 | Always accepts corrections and tries to comply with them | 0.47 | 0.85 |
| 136 | 21 | Willingly allows others to take the lead | 0.39 | 0.67 |

III. Ability to communicate through Speech.

| Check-list no. | Serial no. | Wording of item | $r_{pbi}$ | $p$ value |
|---|---|---|---|---|
| 111 | 1 | Has no difficulties in expressing himself or herself orally | 0.83 | 0.68 |
| 107 | 2 | Has an indistinct and defective pronounciation | 0.83 | 0.21 |
| 102 | 3 | Speaks in monosyllables and has difficulty in forming connected sentences | 0.81 | 0.23 |
| 144 | 4 | Has very few contacts with fellow-workers | 0.75 | 0.30 |
| 126 | 5 | Finds it easy to make contacts with others | 0.74 | 0.77 |
| 139 | 6 | Often walks by himself or herself without talking to his or her fellow-workers | 0.74 | 0.26 |
| 110 | 7 | Is very retiring | 0.50 | 0.14 |
| 47 | 8 | Avoids contact with his or her fellow-workers | 0.38 | 0.08 |
| 134 | 9 | Is barely tolerated by his or her fellow-workers | 0.37 | 0.06 |
| 106 | 10 | Communicates with others only by signs and gestures | 0.31 | 0.05 |
| 145 | 11 | Often uses words wrongly | 0.19 | 0.14 |

IV. Motivation to Work

| Check-list no. | Serial no. | Wording of item | $r_{pbi}$ | $p$ value |
|---|---|---|---|---|
| 44 | 1 | Works fitfully and takes frequent breaks | 0.85 | 0.35 |
| 119 | 2 | Has satisfactory and settled working habits | 0.83 | 0.64 |
| 87 | 3 | Works at an even rate, calmly and systematically | 0.82 | 0.59 |
| 1 | 4 | Sticks to his or her work and works without unnecessary delays | 0.82 | 0.68 |
| 61 | 5 | Keeps up a calm and regular rate of work the whole day | 0.81 | 0.62 |
| 120 | 6 | Works without needing to be urged on | 0.79 | 0.67 |
| 66 | 7 | Begins work in good time and is not in a hurry to leave off | 0.77 | 0.53 |
| 13 | 8 | Works perseveringly, if he or she is not disturbed by others | 0.75 | 0.76 |
| 75 | 9 | Works well periodically but is inactive between the periods | 0.73 | 0.42 |
| 59 | 10 | Takes the work seriously and feels responsible for it | 0.72 | 0.80 |
| 77 | 11 | Quite often lengthens the breaks by 10-15 minutes | 0.71 | 0.30 |
| 100 | 12 | Likes to work in a concentrated fashion | 0.71 | 0.67 |
| 25 | 13 | Spends too much time in corridors and toilets | 0.70 | 0.24 |
| 63 | 14 | Starts work at once after a break | 0.70 | 0.71 |
| 86 | 15 | Is completely absorbed in his work | 0.70 | 0.42 |
| 104 | 16 | Works assiduously and purposefully | 0.70 | 0.44 |
| 141 | 17 | Is very industrious | 0.70 | 0.45 |
| 14 | 18 | Requires constant supervision to make him or her sit still att work | 0.69 | 0.17 |
| 23 | 19 | Has a consistently positive attitude to the work | 0.68 | 0.64 |
| 140 | 20 | Works unevenly from one day to another | 0.68 | 0.42 |
| 35 | 21 | Is always punctual | 0.67 | 0.50 |
| 78 | 22 | Easily gives up in setbacks at work | 0.67 | 0.38 |
| 2 | 23 | Often breaks off work to talk with fellow-workers | 0.66 | 0.36 |
| 103 | 24 | Is anxious to complete tasks that he has begun | 0.66 | 0.74 |
| 89 | 25 | Is anxious to make a good effort | 0.61 | 0.77 |
| 18 | 26 | Seldom gives the impression of being tired during working hours | 0.57 | 0.71 |
| 3 | 27 | Shows signs of enjoying the work | 0.57 | 0.85 |
| 76 | 28 | Tires quite quickly at monotonous work and wants a change | 0.56 | 0.29 |
| 41 | 29 | Is anxious to leave everything in order at the workplace | 0.54 | 0.52 |
| 21 | 30 | Has shown a more negative attitude to work as time has gone by | 0.54 | 0.14 |
| 11 | 31 | Complains that the work is monotonous | 0.51 | 0.21 |
| 80 | 32 | Gives up completely when the work seems to be too difficult for him or her | 0.33 | 0.35 |
| 71 | 33 | Is never absent from work without a proper reason | 0.29 | 0.85 |
| 22 | 34 | Has shown a more positive attitude to work as time has gone by | 0.27 | 0.45 |

V. Movement Co-ordination and Working Pace.

| Check-list no. | Serial no. | Wording of item | $r_{pbi}$ | $p$ value |
|---|---|---|---|---|
| 148 | 1 | Works with sure and well-balanced movements | 0.82 | 0.76 |
| 149 | 2 | Has great difficulties with tasks which require handiness | 0.79 | 0.29 |
| 155 | 3 | Has comparatively quick and flexible fingers | 0.76 | 0.67 |
| 151 | 4 | Handles tools worse than most | 0.76 | 0.17 |
| 42 | 5 | Works in a fumbling and awkward manner | 0.73 | 0.14 |
| 131 | 6 | Familiarizes himself quite quickly with new tasks | 0.73 | 0.85 |
| 24 | 7 | Requires unusually detailed instructions to master a new task | 0.71 | 0.26 |
| 125 | 8 | When he or she works, he or she does so at a markedly slow rate | 0.70 | 0.26 |
| 95 | 9 | Is unusually slow in all his or her movements | 0.70 | 0.18 |
| 114 | 10 | Is comparatively quick at his or her work | 0.63 | 0.55 |
| 150 | 11 | Has firm and steady hands | 0.61 | 0.86 |
| 81 | 12 | Has stiff and rather unflexible fingers | 0.61 | 0.11 |
| 16 | 13 | Sometimes complains that the work is difficult | 0.59 | 0.35 |
| 57 | 14 | Is inactive for long periods | 0.57 | 0.35 |
| 54 | 15 | Is anxious and wary at the prospect of a new task | 0.49 | 0.33 |
| 108 | 16 | Often works with jerky movements | 0.46 | 0.15 |
| 90 | 17 | His or her performance in the workshop is average | 0.43 | 0.74 |
| 123 | 18 | Will set about new tasks without delay, as soon as he or she has finished a certain assignment | 0.43 | 0.68 |
| 39 | 19 | Is uncertain and underestimates his or her ability | 0.43 | 0.23 |
| 19 | 20 | Is worried lest he or she should make a misktake | 0.38 | 0.32 |

VI. Technical Knowledge and Skills.

| Check-list no. | Serial no. | Wording of item | $r_{pbi}$ | $p$ value |
|---|---|---|---|---|
| 154 | 1 | Can take over other workers' tasks in the workshop without much difficulty | 0.75 | 0.53 |
| 112 | 2 | Has no objection to making a little effort to finish an urgent task | 0.70 | 0.59 |
| 105 | 3 | Can be employed on inspection work of different kinds | 0.67 | 0.41 |
| 91 | 4 | Is very willing to learn more about processes, machines and tools | 0.64 | 0.39 |
| 36 | 5 | Has apparently no ambition to increase his or her working capacity | 0.64 | 0.41 |
| 143 | 6 | Is quite unable to teach his or her fellow-workers | 0.64 | 0.38 |
| 96 | 7 | Does not care about the quality of the work | 0.63 | 0.18 |
| 55 | 8 | Helps his or her fellow-workers on his or her own initiative, whenever necessary | 0.63 | 0.62 |
| 52 | 9 | Works independently and does not need much guidance | 0.60 | 0.79 |
| 46 | 10 | Has no understanding of the quality of the work done | 0.60 | 0.17 |
| 30 | 11 | Contributes quantitatively more than most to the workshop's production | 0.59 | 0.35 |
| 94 | 12 | Has a fairly realistic idea of his or her own capacity | 0.59 | 0.47 |
| 147 | 13 | Is very attentive and interested when he or she is given instructions | 0.58 | 0.86 |
| 121 | 14 | Has generally to be responsible for more than his or her own share of a team work job | 0.57 | 0.27 |
| 64 | 15 | Is interested in the results of his or her own work | 0.57 | 0.77 |
| 98 | 16 | Is unwilling to accept responsibility for occasional tasks | 0.55 | 0.32 |
| 28 | 17 | Often drops material and tools on the floor | 0.49 | 0.11 |
| 62 | 18 | Is careless as regards cloths and personal appearance | 0.49 | 0.15 |
| 124 | 19 | Knows very well what he or she wants | 0.48 | 0.65 |
| 146 | 20 | May show qualities of leadership in certain tasks | 0.45 | 0.38 |
| 99 | 21 | Follows carefully the working instructions given | 0.44 | 0.91 |
| 65 | 22 | Consistently gives poor performances from the point of view of quality | 0.37 | 0.05 |
| 117 | 23 | Generally puts things down in a calm and well-poised manner | 0.32 | 0.91 |
| 40 | 24 | Puts quite penetrating questions about work processes and working methods | 0.26 | 0.18 |

*Appendix B:2.* Sub-Scales in a Five-Degree Rating Schedule. $n = 66$ (40 men and 26 women)

I Rated work proficiency.

| Schedule no. | Serial no. | Wording of item | $r_{pbi}$ | Mean value |
|---|---|---|---|---|
| 7 | 1 | Ability to understand instructions | .85 | 3.45 |
| 13 | 2 | Initiative in work situations | .84 | 3.18 |
| 14 | 3 | Ability to carry out tasks without supervision | .76 | 3.18 |
| 6 | 4 | Self-confidence in work situations | .61 | 3.17 |
| 8 | 5 | Quality of work | .60 | 3.44 |
| 10 | 6 | Deftness (rapidity of movement and manipulation with hands and fingers) | .54 | 3.17 |
| II Rated adjustment to work | | | | |
| 5 | 1 | Persistence at work | .71 | 3.38 |
| 1 | 2 | Interest in work | .66 | 3.30 |
| 3 | 3 | Orderliness at place of work | .63 | 3.18 |
| 4 | 4 | Behaviour and personal style | .57 | 3.21 |
| 12 | 5 | Ability to co-operate with supervisors | .52 | 3.24 |
| 11 | 6 | Ability to co-operate with fellow-workers | .48 | 2.98 |
| 2 | 7 | Punctuality | .48 | 3.20 |

# Appendix C

# Detailed descriptions of eight jobs

## 1. Pressing eyelets

*Description of workplace and cycle of operations.* The job has a short cycle and is comparatively simple. It is performed on an excenter press of Arosverken make (type AL 10). A brass blank, stamped out in advance, is placed in a pre-determined position and in a pre-determined way in a bending tool, after which the press is operated by both hands and the bending takes place.

Some blanks may be faulty and in that case must be scrapped. On account of the coating of oil, two blanks may sometimes stick together and in these cases they have to be separated. The operator must also take care to place the blank correctly in the bending tool and see that the characteristic lug on one side of the eyelet is always correctly placed.

The work is to some extent oily but does not require any special protective clothing. However, the operator's ears should be protected against the noise of the press. The operator works sitting down and the height from the floor is 82 cm. Cf. Figs. C2 and C3 on page 188.

According to the expert in methods-time measurement (MTM), the task makes great demands on the operator's persistence, on account of the very short and monotonous work cycle. On the other hand, the demands are moderate as regards psychomotor skill, speed of movement and concentration and small as regards physical strength. The period of training for a normal worker is estimated at 5 weeks and the operator should be familiar with the job after a further 3 weeks. After 4 days, a normal worker should be able to achieve 90 % of the MTM output. The pressing of an eyelet is reckoned to take 106.6 TMU or 3.8 seconds. The operator who achieves this rate will consequently produce 844 details per working hour. (As in the other tasks, the calculation in this case was based on the TMU value and the assumption that the effective working time is 54 minutes per hour.)

*Evaluation of performances.* The scale of the production is indicated by a counter, which records the number of movements made by the press. In addition, the production in each working period is checked by weighing and inspected, when eyelets which have been wrongly placed in the bending tool and are consequently faulty are scrapped.

*Report on performances and development of performances.* In this task, the quantitative measurement of working capacity is expressed as the number of approved eyelets pressed per hour during the last 2 working days. The distribution of these measurements between men and women is shown in Table C1, which reveals that the task has a very good ability to differentiate.

The development of the performances is shown in Fig. C1, from which the numbers of approved eyelets pressed per hour during eight successive working periods can be read off. The graphs for both men and women clearly show the effects of training. The men consistently reach a somewhat higher level of output. If we disregard the first day, the performances of both sexes are better during the mornings than during the afternoons.

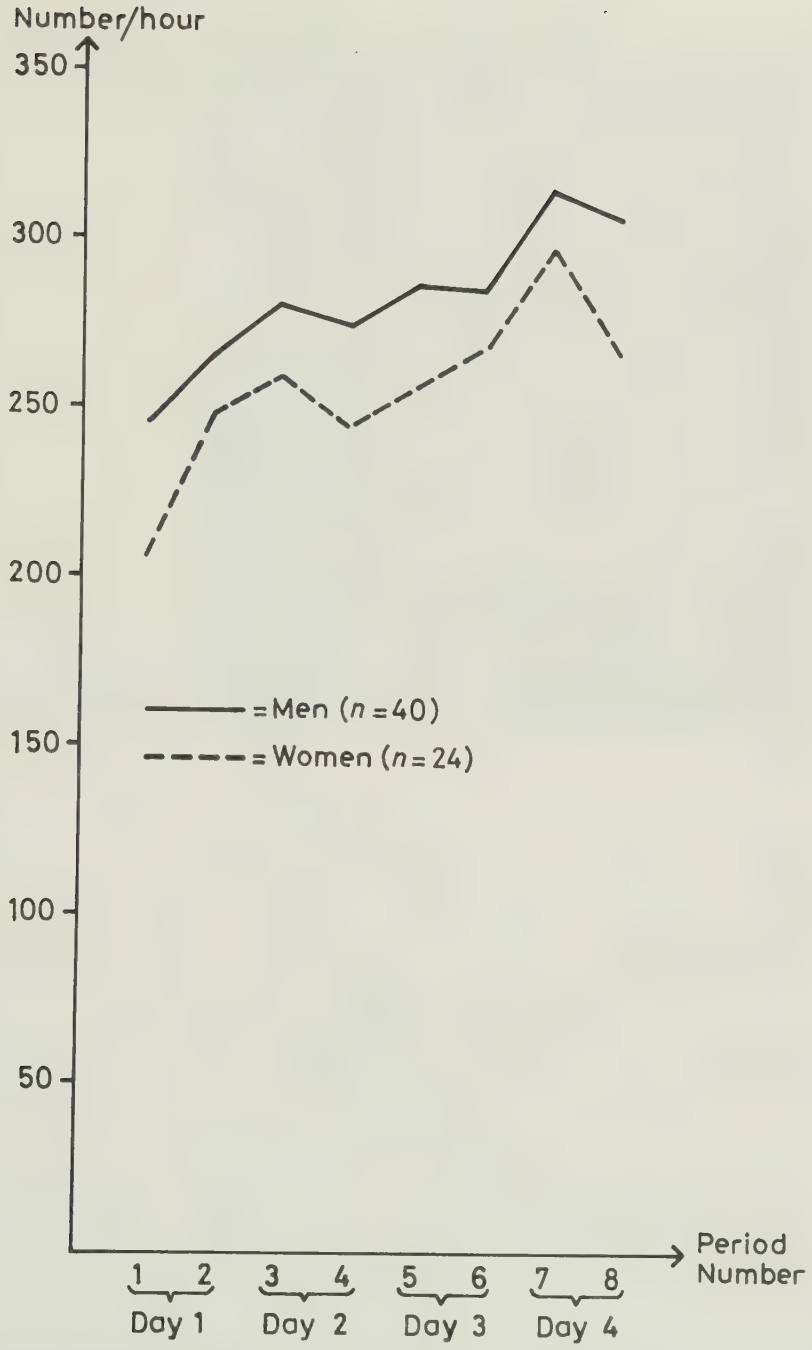



Fig. C 1. Development of performance in the pressing of eyelets.

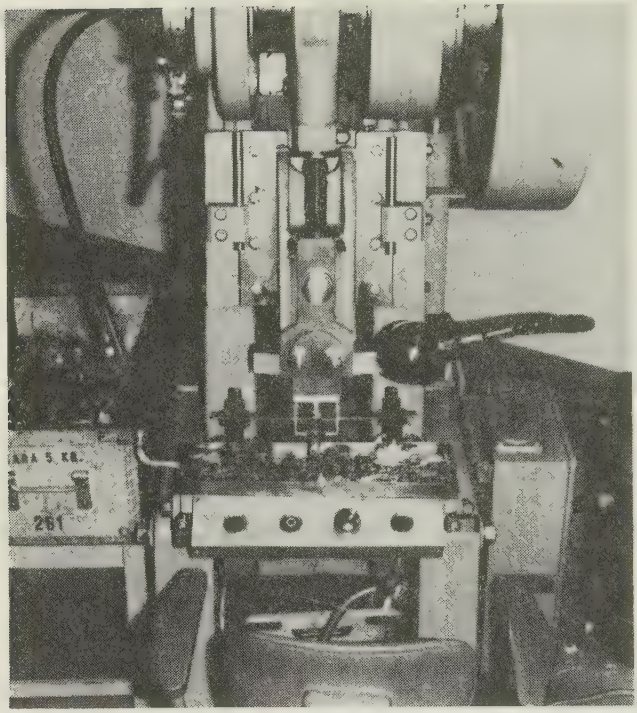


Fig. C 2. Photo of workplace for pressing eyelets.

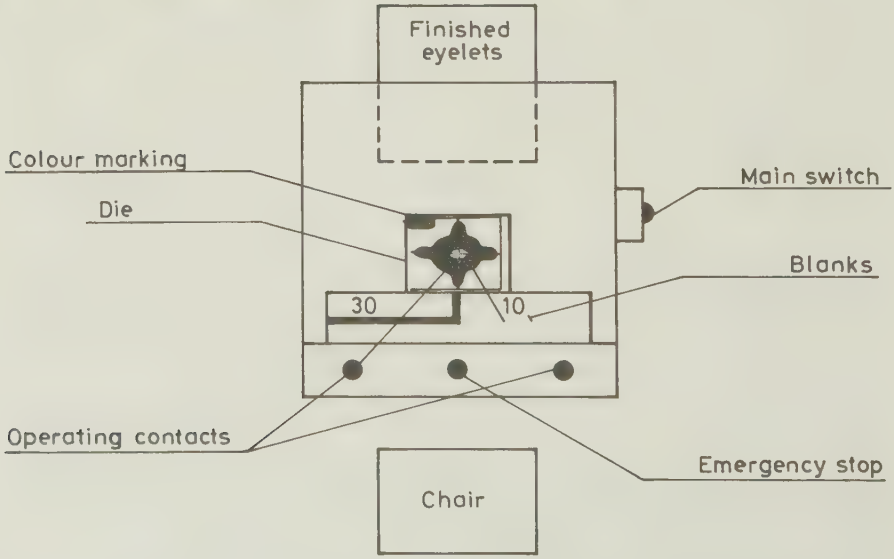



Fig. C 3. Sketch plan of workplace for pressing eyelets.

*Table C 1.* Number of eyelets pressed per hour.

| Details per hour | Men | Women |
|---|---|---|
| 0 – 49 | | |
| 50 – 99 | 1 | 1 |
| 100 – 149 | 4 | 1 |
| 150 – 199 | 6 | 3 |
| 200 – 249 | 6 | 6 |
| 150 – 299 | 6 | 2 |
| 300 – 349 | 6 | 6 |
| 350 – 399 | 2 | 1 |
| 400 – 449 | 2 | 2 |
| 450 – 499 | 2 | 1 |
| 500 – 549 | 1 | 1 |
| 550 – 599 | 4 | |
| Σ | 40 | 24 |
| M | 295 | 278 |
| S.D. | 138 | 103 |

## 2. Cutting threads in nuts

*Description of workplace and cycle of operations.* The task involves the manual feeding forward of a circular-feed magazine for nut blanks, by means of a lever, after which the machine is set in motion by a pedal and does the threading, while simultaneously a fresh nut blank is placed in the magazine.

Some of the nut blanks may be faulty and must in that case be scrapped. Obviously, only one blank may be placed in position at each feeding movement. Furthermore, the blanks must be correctly placed in the magazine. They must not come out askew and must lie flat on the bottom of the magazine. It is also necessary that the magazine should be properly fed forward. The feeding movement must come full butt against a stop in both directions.

The task has a short work cycle and is highly mechanized. The threading machine is of the Hüller make (type UG-3) and the working height is 110 cm.

The work is done with the operator seated and is described as oily, so that a special protective apron must be worn. In addition, the task can be adjusted for both right- and left-handed workers, in that the magazine has feeding levers on both the right-hand and the left-hand sides. Cf. Figs. C5 and C6 on page 191.

According to the MTM expert, the task makes comparatively moderate demands on psychomotor skill, concentration and persistence, while the requirements for speed of movement and physical strength are smaller. The period of training for a normal worker is assessed as short and it should be possible to describe the worker as used to the job after only 2 days. The threading of a nut is reckoned to take 275.4 TMU or 9.9 seconds. The operator who achieves this rate will consequently produce 327 details per working hour.

*Evaluation of performances.* The production is recorded by two counters. One records the number of working movements made by the machine, while the other records the

number of times that the magazine is fed forward. The production in each working period is checked by weighing and inspected, when any unthreaded nuts are discarded.

*Report on performances and development of performances.* The quantitative measurement of working capacity in this task is expressed by the number of correctly threaded nuts produced per hour during the last 2 working days. The distribution, average value and variation of this measurement in men and women are shown in Table C2. The task differentiates well and the distributions seem to approach the shape of the normal distribution.

The development of performances is shown in Fig. C4, which gives the number of correctly threaded nuts produced per hour during eight consecutive working periods. The form of the diagram indicates that there was a certain effect of training in both men and women. It is also evident that the morning period yields a somewhat greater output than the afternoon period. The men's performances are consistently better than the women's.

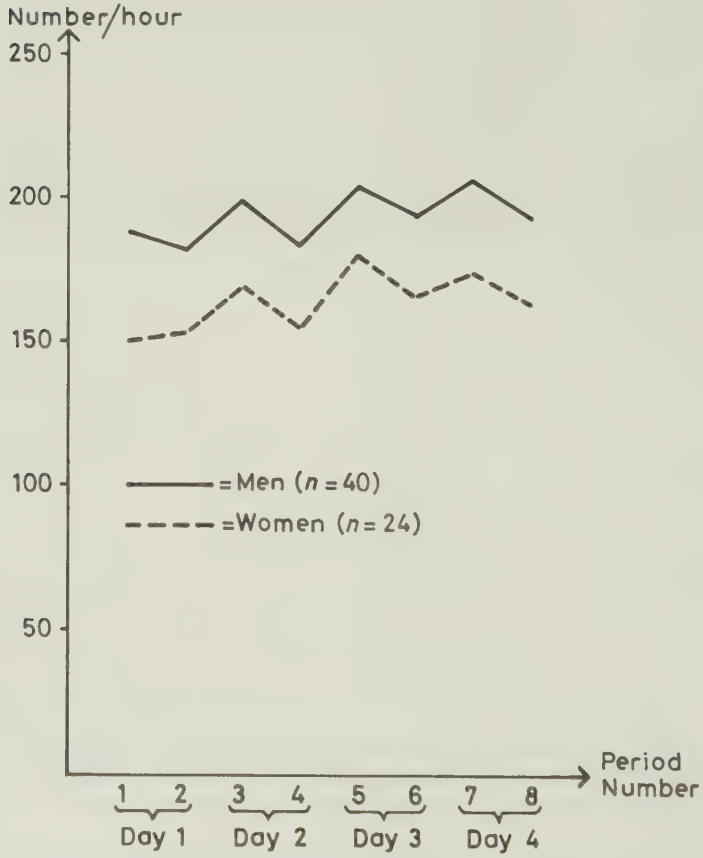



Fig. C 4. Development of performance in threading nuts.

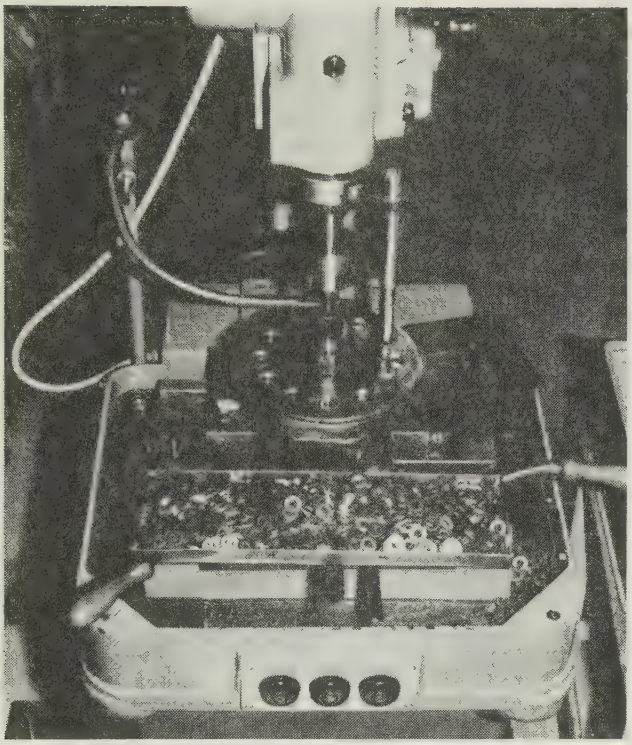


Fig. C 5. Photo of workplace for threading nuts.

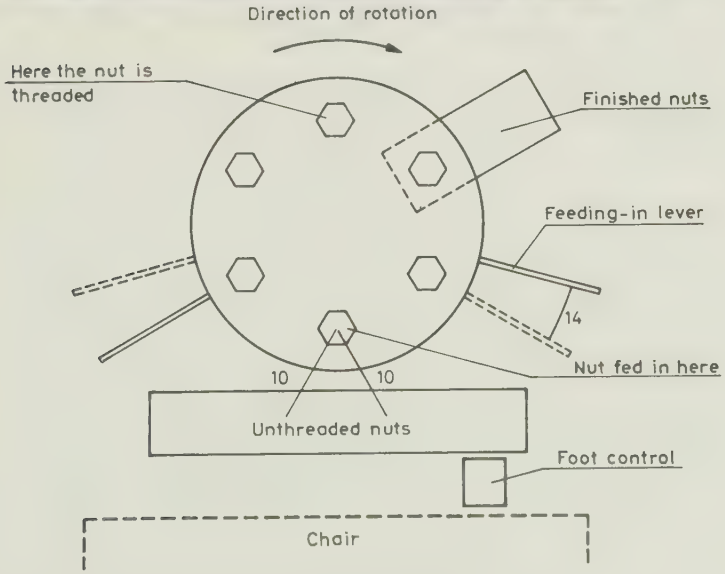



Fig. C 6. Sketch plan of workplace for threading nuts.

*Table C 2.* Number of nuts threaded per hour.

| Details per hour | | Men | Women |
|---|---|---|---|
| 60 – 79 | | 1 | 1 |
| 80 – 99 | | 3 | 3 |
| 100 – 119 | | 2 | 3 |
| 120 – 139 | | 2 | 1 |
| 140 – 159 | | 5 | 1 |
| 160 – 179 | | 3 | 4 |
| 180 – 199 | | 2 | 3 |
| 200 – 219 | | 6 | 2 |
| 220 – 239 | | 6 | 3 |
| 240 – 259 | | 1 | 1 |
| 260 – 279 | | 3 | 1 |
| 280 – 299 | | 3 | 1 |
| 300 – 319 | | 1 | |
| 320 – 339 | | 1 | |
| 340 – 359 | | | |
| 360 – 379 | | 1 | |
| | Σ | 40 | 24 |
| | M | 199 | 171 |
| | S.D. | 71 | 61 |

## 3. Turning holders

*Description of workplace and cycle of operations.* The task involves the drilling, turning and final checking of a brass blank. A centre drill, fed forward by hand, is used for the drilling, and the turning along is carried out semi-automatically, i.e. by the manual engagement of a sliding feed and fine adjustment against a stop in the concluding phase. The final check is made with a snap gauge. The work is described as being fairly dirty and is done standing. The lathe is of the Blomqvist make (type DS 600 K Master) and has a working height of 110 cm. Cf. Figs. C8 and C9 on page 195.

The worker has first to determine which end of the blank is chamfered and then to insert it into the clamping sleeve against a stop, whereupon the blank is firmly locked. It is also clear from the above description that the feeding movements in both the drilling and the concluding phase of the turning along are made manually against a stop. In the final check in the snap gauge, the worker has to decide whether the holder has the intended dimension or not.

According to the MTM expert, the task makes moderate demands on psychomotor ability, strength, ability to co-ordinate and persistence. The requirement for concentration is described as great, while the requirement for speed of movement is considered to be small. The period of training for a normal worker is estimated to be 1 week. After another week, it should be possible to describe the worker as used to the job. According to the MTM analysis, the whole cycle of operations is calculated to take 760.3 TMU or 27.3 seconds. The worker who achieves this rate will consequently produce 118 details per working hour.

*Evaluation of performances.* The production in each working period is checked by counting. Each detail is also checked by measuring the depth of the drilled hole, the length of the turning and the turned diameter by means of a dial test indicator and a snap gauge. Details which do not meet the given tolerances, in accordance with the drawing, are rejected. A note is also made of the type of fault which gave rise to the rejection.

*Report on performances and development of performances.* From the beginning, we did not consider that a thorough check was necessary but contented ourselves with spot checks. This meant that we did not obtain any satisfactory measurement of the working performances of the first 10 experimental subjects. In order to make up for this deficiency, we had them work for a fifth day somewhat later on. As far as they were concerned, the measurement of working capacity was then calculated as the number of approved holders turned per hour during this fifth working day.

As far as the other workers were concerned, the regular procedure was naturally applied. Thus, the quantitative measurement of working capacity is expressed by the number of details turned per hour during the last 2 working days. The distribution, mean value and variation of this measurement may be seen in Table C3. As previously, we find in this case that the task has a satisfactory ability to differentiate. The distributions seem to approach the shape of the normal distribution, though with a certain tendency to positive skewness among the women.

In consequence of the above-mentioned deficiency in the evaluation technique at the beginning of the investigation, the development of the performances could be calculated for only 30 men and 18 women. It is shown in Fig. C7. Training effects can be clearly traced in this task also and it is evident that the men produce considerably more than the women.

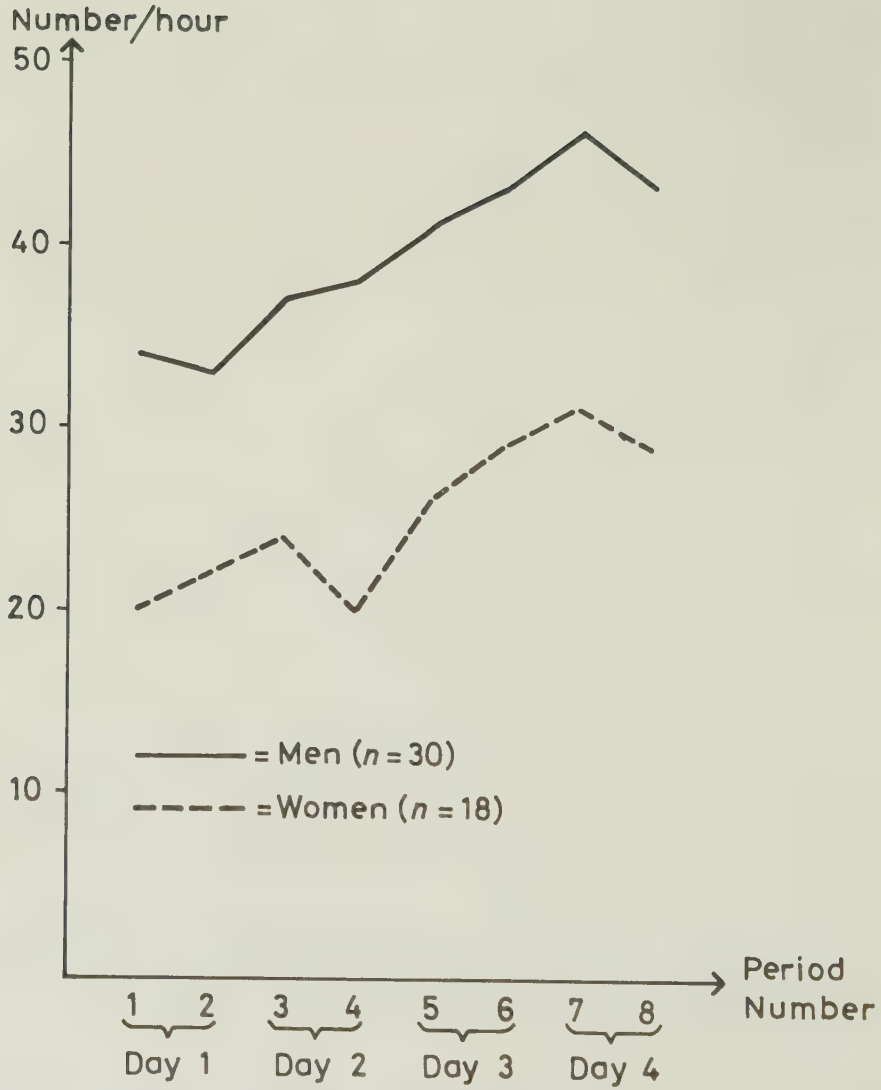



Fig. C 7. Development of performance in turning holders.

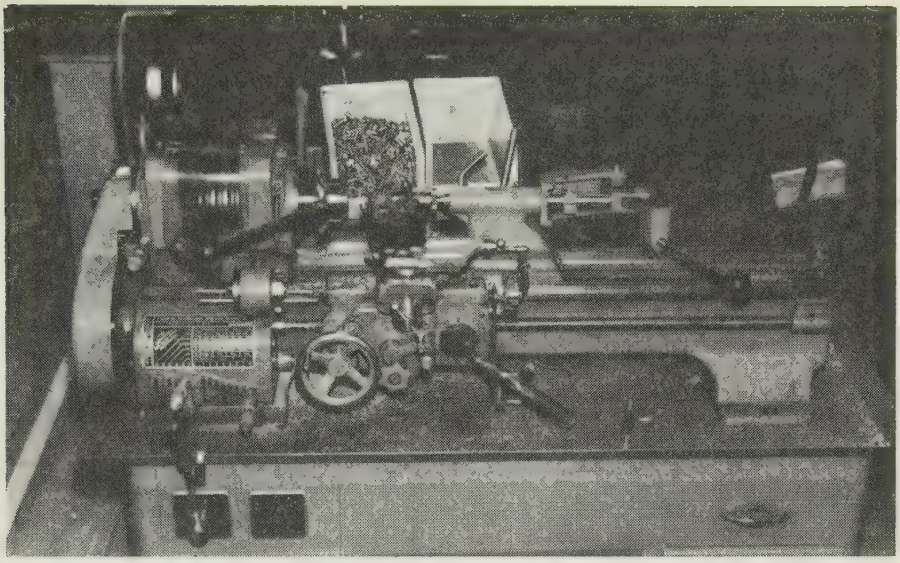


Fig. C 8. Photo of workplace for turning holders.

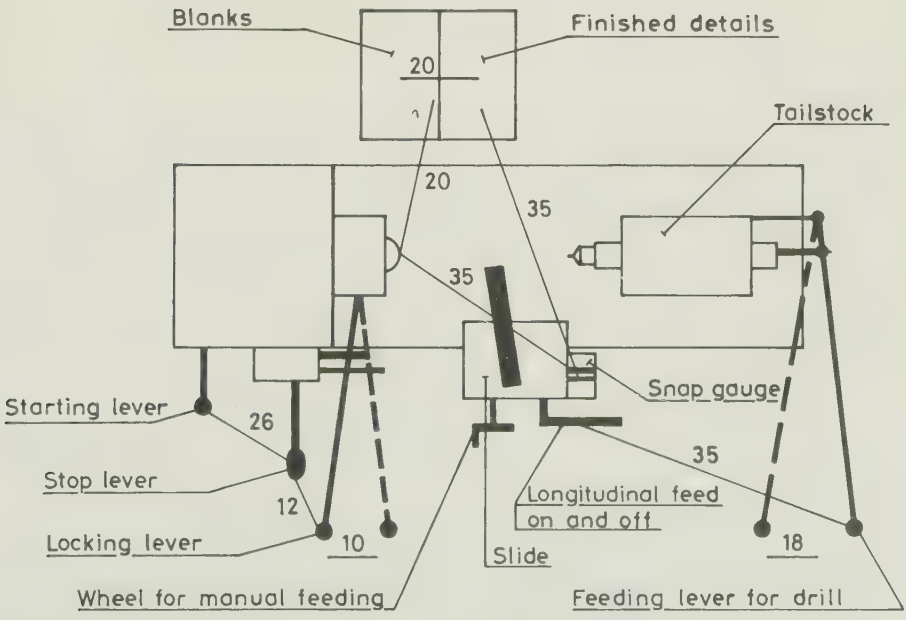



Fig. C 9. Sketch plan of workplace for turning holders.

*Table C 3.* Number of holders turned per hour.

| Details per hour | Men | Women |
|---|---|---|
| 5 – 9 | 1 | 3 |
| 10 – 14 | 4 | 3 |
| 15 – 19 | 1 | 2 |
| 20 – 24 | 3 | 2 |
| 25 – 29 | 4 | 2 |
| 30 – 34 | 1 | 1 |
| 35 – 39 | 3 | 3 |
| 40 – 44 | 5 | 4 |
| 45 – 49 | 3 | 1 |
| 50 – 54 | 1 | 1 |
| 55 – 59 | 3 | 1 |
| 60 – 64 | 5 | |
| 65 – 69 | 1 | 1 |
| 70 – 74 | 2 | |
| 75 – 79 | 1 | |
| 80 – 84 | 1 | |
| 85 – 89 | | |
| 90 – 94 | 1 | |
| 95 – 99 | | |
| Σ | 40 | 24 |
| M | 43.6 | 29.7 |
| S.D. | 21.52 | 16.63 |

## 4. Assembly of brackets

*Description of workplace and cycle of operations.* The task involves the simple assembly of a screw, two plates and a nut. All the details are made of steel. The work is carried out sitting at a table, which is 79 cm high, 80 cm wide and 160 cm long. The details are kept in four plastic boxes for unassembled brackets, screws, plates and nuts separately.

In order to facilitate the assembly work, the bracket is placed in a beechwood work fixture, the external dimensions of which are 170 x 55 x 50 cm. The plastic boxes are placed in approximately a semicircle round the fixture. The finished bracket is dropped into a channel and passes a counter, which records the number of details assembled. Cf. Figs. C11 and C12 on page 198.

As in the other tasks, the worker has to learn the correct working sequence and the different details in the procedure. The requirement that the nut shall be properly tightened may cause difficulties here. The worker must also be able to see if any of the details is damaged and set it aside.

According to the MTM expert, the task makes moderate demands on psychomotor ability, speed of movement, strength and ability to co-ordinate, while the requirement for concentration is described as small. On the other hand, the worker's persistence must be good, on account of the monotonous character of the work. The training or learning period for a normal worker is considered to be short and normal production should be achieved after about 3 days, according to the MTM expert. After a further 3 days, it

should be possible to describe the worker as being used to the job. In the MTM analysis, the assembly of a bracket is calculated to take 307.1 TMU or 11.05 seconds. The worker who achieves this rate will consequently produce 293 details per working hour.

*Evaluation of performances.* The production in each working period is noted from the counter and inspected, when the brackets which do not meet the stipulated requirements are rejected. Faults which commonly occur are that the nut has not been properly tightened—which, in its turn, may be connected with the fact that the worker has not rejected a defective detail—or that one of the two plates has not been assembled.

*Report on performances and development of performances.* The quantitative measurement of working capacity is accordingly calculated as the number of approved brackets assembled per hour during the last 2 working days. The distribution, mean value and variation of this measurement are shown in Table C4. This task also has a satisfactory ability to differentiate. The shape of the distribution displays a certain tendency to positive skewness in both men and women. The men have, on the average, a somewhat higher level of performance.

The development of the performances during the 4 working days is shown in Fig. C10. As far as the men are concerned, the curve is unusually even, with signs of a moderate training effect. As regards the women, the curve shows a somewhat more uneven course, with a slighter training effect.

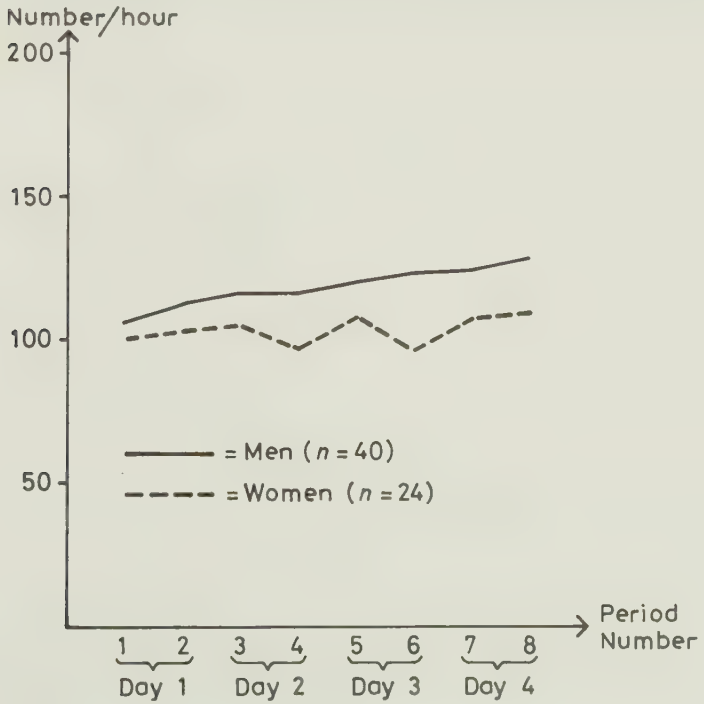



Fig. C 10. Development of performance in assembling brackets.

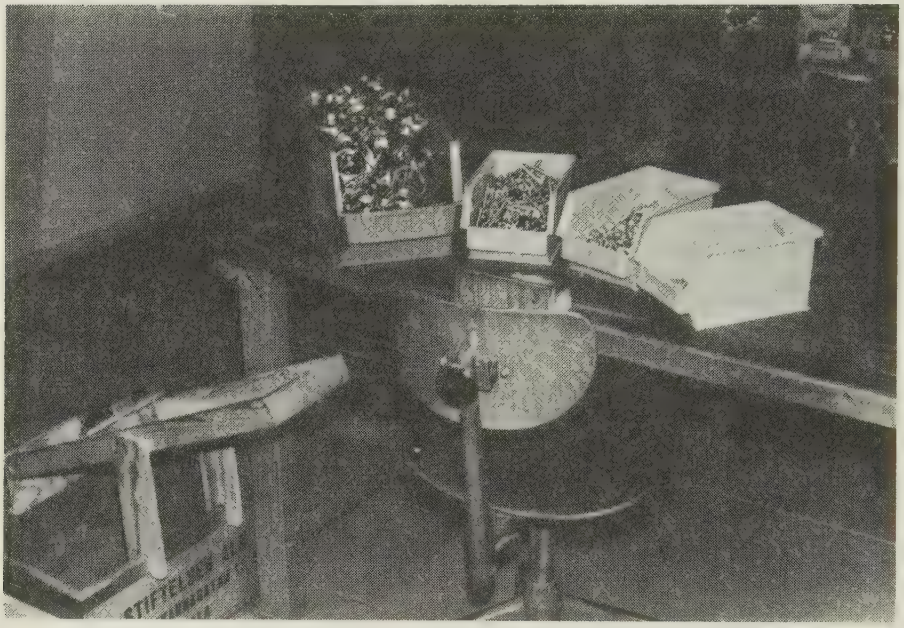


Fig. C 11. Photo of workplace for assembling brackets.

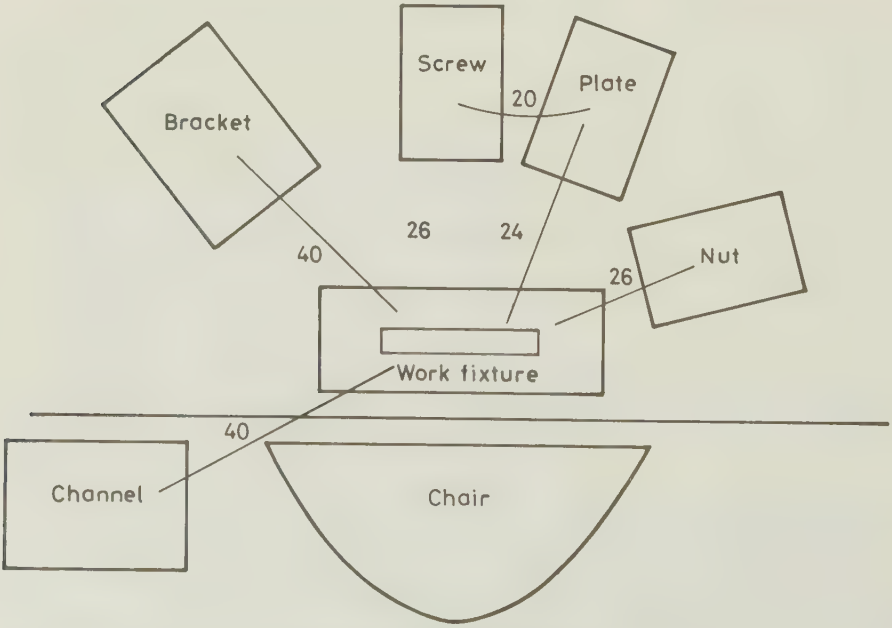



Fig. C 12. Sketch plan of workplace for assembling brackets.

*Table C 4.* Number of brackets assembled per hour.

| Details per hour | Men | Women |
|---|---|---|
| 40 – 49 | 2 | 2 |
| 50 – 59 | | 1 |
| 60 – 69 | 2 | 2 |
| 70 – 79 | 4 | 1 |
| 80 – 89 | 3 | 6 |
| 90 – 99 | 4 | 1 |
| 100 – 109 | 3 | |
| 110 – 119 | 3 | 3 |
| 120 – 129 | 3 | 2 |
| 130 – 139 | 2 | |
| 140 – 149 | 3 | 2 |
| 150 – 159 | 3 | 1 |
| 160 – 169 | | 1 |
| 170 – 179 | 2 | 2 |
| 180 – 189 | 2 | |
| 190 – 199 | 1 | |
| 200 – 209 | | |
| 210 – 219 | 1 | |
| 220 – 229 | 1 | |
| 230 – 239 | | |
| 240 – 249 | | |
| 250 – 259 | | |
| 260 – 269 | | |
| 270 – 279 | 1 | |
| Σ | 40 | 24 |
| M | 124 | 105 |
| S.D. | 50.7 | 38.4 |

# 5. Assembly of electrical panels

*Description of workplace and cycle of operations.* The task involves mounting on a specially made board, measuring 15 x 30 cm, a lamp-holder, a switch, a connection box and a plug and connecting these components together with lengths of flex. A pattern is set up in front of the worker, in which he can see how the lengths of flex are to be run. The cycle of operations is a long one and includes shaping, stripping and bending the flex. The work is carried out sitting at a table measuring 160 x 80 x 80 cm. In order to facilitate the work, the table-top is tilted about 10°. All the materials are kept in plastic boxes within comfortable reach of the worker. Cf. Figs. C14 and C15 on page 202.

In order to master the task, the worker must learn the rather long sequence of movements and be able to carry out satisfactorily all the comparatively heterogeneous operations which the task consists of. Some of these operations require him to make a choice between two alternatives or to adapt and differentiate his movements with reference to the existing situation. Thus, he must be able to position the switch correctly, i.e. with the contact screws facing the lamp socket. He must know how to connect a green flex to a green terminal and a red flex to a red terminal. He must be able to attach the flexes properly under the screw heads and to screw up all the screws hard. Finally, it is important that the pin in the plug should be screwed down far enough to fix the flex in the correct position.

The worker himself checks his work, after each panel has been completed, with the aid of an electric lamp and a bell transformer (5 V) as a current source. If the different operations have been carried out correctly, the lamp will light up, which is a stimulating experience for the worker.

According to the MTM expert, the task makes comparatively great demands on psychomotor ability, speed of movement, concentration and ability to use tools, while the requirement for persistence is considered to be smaller.

The necessary training period is considered to be about a month. The worker is considered to be used to the job only after another month. After so short a time as 4 days, the production of the normal worker is estimated to amount to 60 % of the MTM value. According to the MTM analysis, the assembly of an electrical panel requires 7269.2 TMU or 4.36 minutes. The worker who achieves this rate will consequently produce 12.4 details per working hour.

*Evaluation of performances.* The production in each working period is noted and inspected. The examination of the first electrical panels assembled showed that five types of intolerable faults occurred with a certain frequency, viz.:

1. Pin in plug not screwed down far enough.
2. Wire not properly fastened under the screw head in the lamp-holder.
3. Screw not tightened.
4. Cover of connection box not properly fitted.
5. Wrong screw used.

Panels with one or more of these faults are rejected and the reason for the rejection is given in the record for each worker.

*Report on performances and development of performances.* The measurement of working capacity in this task is calculated as the number of approved electrical panels assembled per hour during the last 2 working days. The distribution mean value and variation of this measurement are shown in Table C5. As far as the men are concerned, the distribution seems to approximate fairly clearly to the shape of the normal curve,

while among the women there appears a certain tendency to positive skewness. The level of the men's performances is higher than that of the women's.

The development of performances during the first 4 working days is shown in Fig. C13. For both men and women, a noticeable training effect is to be seen.

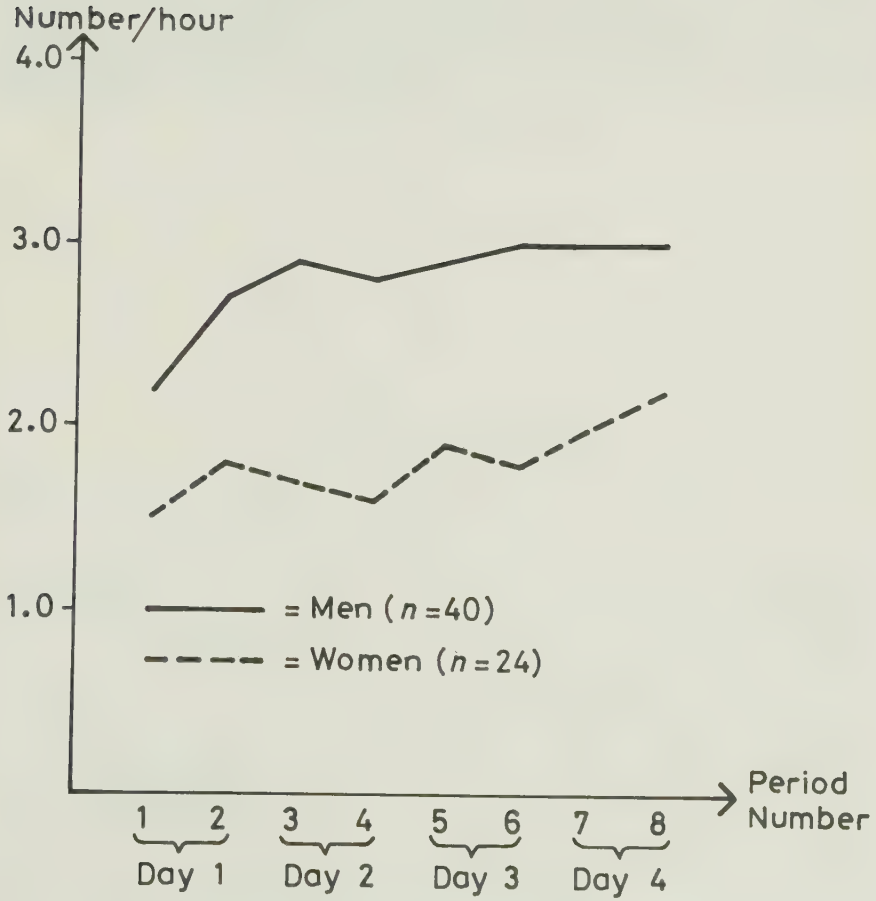



Fig. C 13. Development of performance in assembling electrical panels.

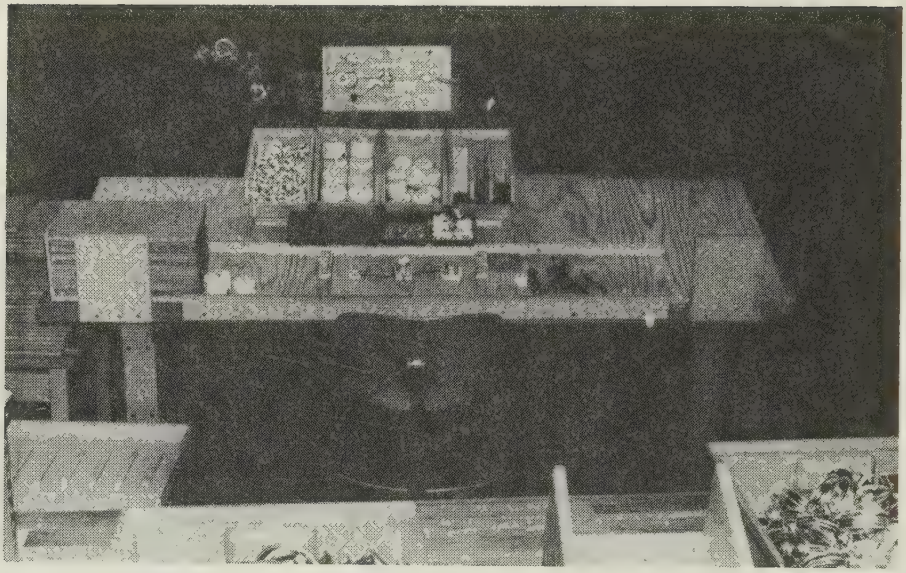


Fig. C 14. Photo of workplace for assembling electrical panels.

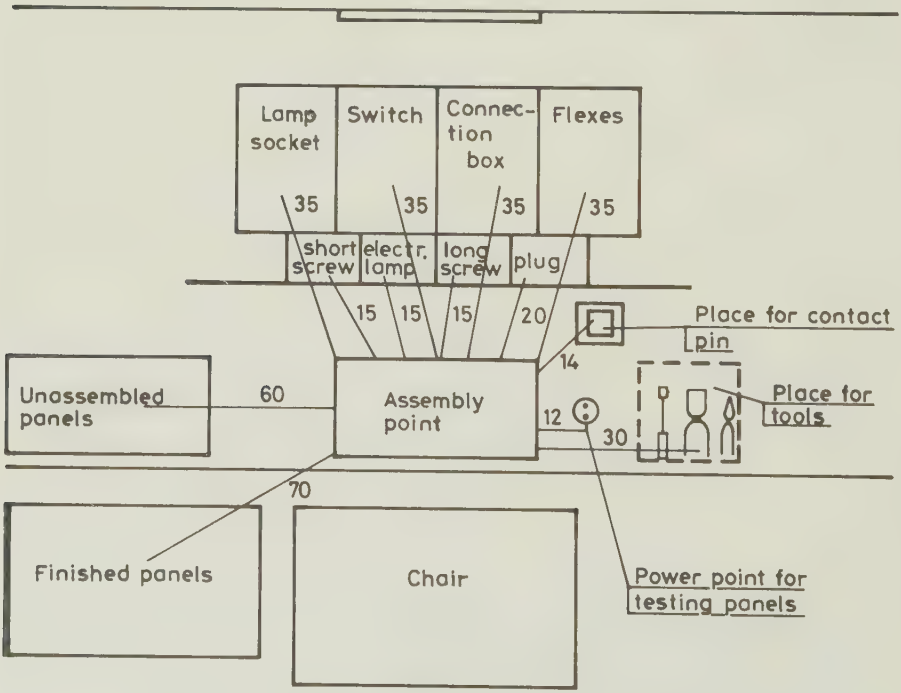



Fig. C 15. Sketch plan of workplace for assembling electrical panels.

*Table C 5.* Number of electrical panels assembled per hour.

| Details per hour | Men | Women |
|---|---|---|
| 0.0 – 0.4 | 2 | 1 |
| 0.5 – 0.9 | 4 | 6 |
| 1.0 – 1.4 | 4 | 6 |
| 1.5 – 1.9 | 2 | 2 |
| 2.0 – 2.4 | 6 | 1 |
| 2.5 – 2.9 | 2 | 2 |
| 3.0 – 3.4 | 5 | 1 |
| 3.5 – 3.9 | 1 | 3 |
| 4.0 – 4.4 | 7 | 2 |
| 4.5 – 4.9 | 3 | |
| 5.0 – 5.4 | | |
| 5.5 – 5.9 | 1 | |
| 6.0 – 6.4 | | |
| 6.5 – 6.9 | 2 | |
| 7.0 – 7.4 | | |
| 7.5 – 7.9 | 1 | |
| 8.0 – 8.4 | | |
| Σ | 40 | 24 |
| M | 3.02 | 2.95 |
| S.D. | 1.82 | 1.25 |

## 6. Making wooden frames

*Description of workplace and cycle of operations.* The task, which is a comparatively complicated one on account of the long and varying cycle of operations, involves the machining and assembling of four wooden blanks cut into lengths beforehand. The work is carried out standing, partly at a band saw and partly at a bench. At the band saw, the trimming of length and breadth, the sawing of dowels and dowel holes, and the oblique bevelling of one edge are carried out with the aid of five wooden patterns. The band saw is of the Kölle make (type B63) and the working height is 97 cm. The patterns run on a movable wooden carriage, specially made for the purpose and placed on the working table of the band saw. The bench work consists of cutting out dowel holes with a wood chisel, planing, assembling, nailing and oblique bevelling of an edge with a plane. All the bench work, except the assembling, is done in work fixtures. Cf. Figs. C17—C20 on pages 206 and 207.

The task presupposes that the worker can learn a comparatively long sequence of movements. Several of the operations also require a certain amount of knowledge and adaptability. Thus, the blanks must be placed in exactly the right positions in the patterns; otherwise there is a risk that play will arise when the pieces are assembled. It is also necessary that the blank shall be pushed in until it reaches the stop in the pattern for the cutting out of the dowel hole, before the wood chisel is used. The blanks must likewise be placed in exactly the right position in the planing pattern. In the assembly work, it is a matter of turning the sawn chamfer outwards, of positioning the frame correctly in the oblique bevelling pattern, and of knocking in the nails in the correct places in the frame.

According to the MTM expert, the task makes moderate demands on psychomotor ability, speed of movement, strength and co-ordination, while the requirements for concentration and ability to follow a pattern of movement are greater. The requirements for concentration and ability to use tools are fairly small. The training period for a normal worker is considered to be 2 months. Another month is reckoned to be needed before the worker becomes familiar with the task. After so short a time as 4 days, the normal worker's production is assessed as being equivalent to 75 % of the MTM value. According to the MTM analysis, the making of a wooden frame is calculated to take 20,510.8 TMU or 12.3 minutes. The worker who achieves this rate will consequently produce 4.4 details per working hour.

*Evaluation of performances.* The production in each working period is noted and inspected. In our experience, the following types of fault occur with a certain frequency:

1. Blank is wrongly positioned in longitudinal sawing.
2. Dowel hole is made too short.
3. Dowel is not hammered right in during assembly.
4. Nail is knocked in obliquely, so that the frame splits.
5. The worker hammers on the frame, so that it splits.

(When faults of types (4) and (5) occur, the wooden frame is rejected if the crack is more than 5 cm long or if it has arisen so near an edge that the piece in question is not firmly anchored to the frame.)

All the faults which occur are recorded in connection with the inspection. Finally, the number of approved details made per unit of time is calculated.

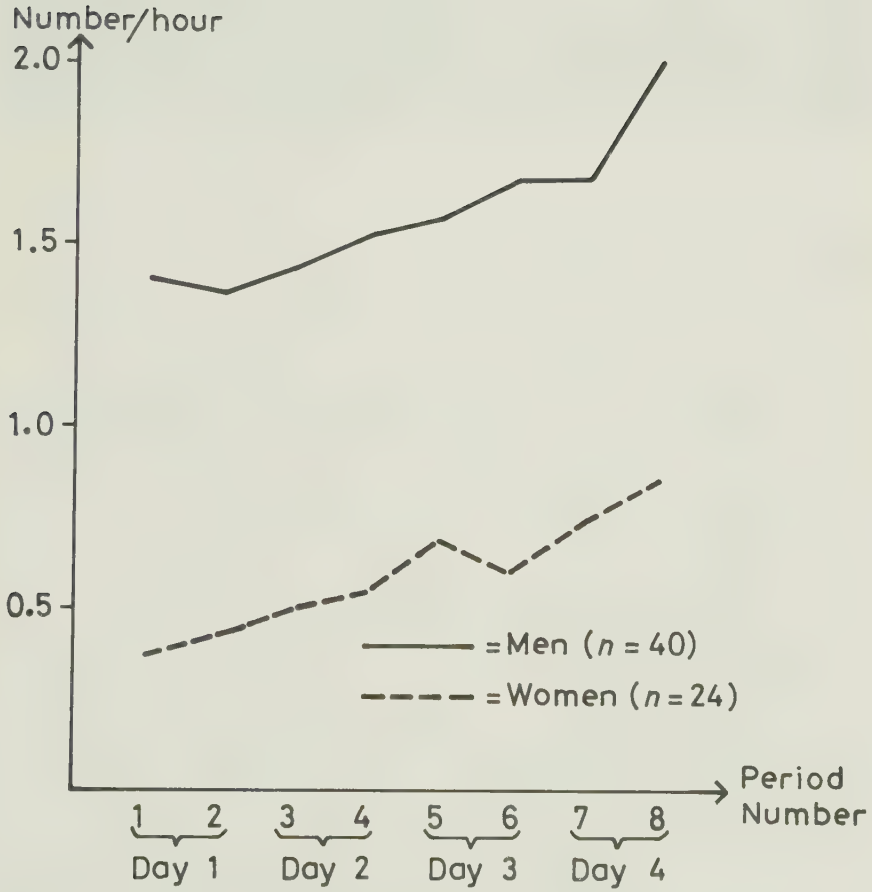



Fig. C 16. Development of performance in making wooden frames.

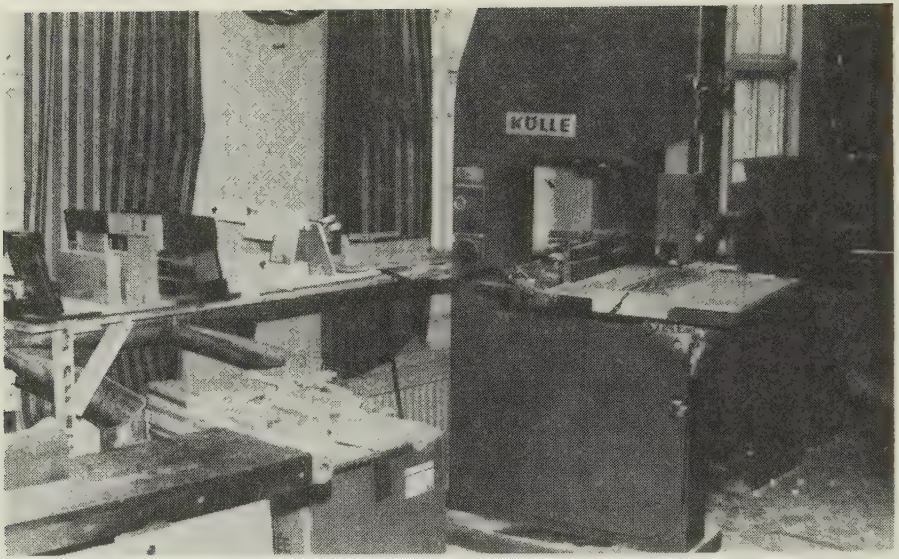



Fig. C 17. Photo of workplace for sawing wooden frames.

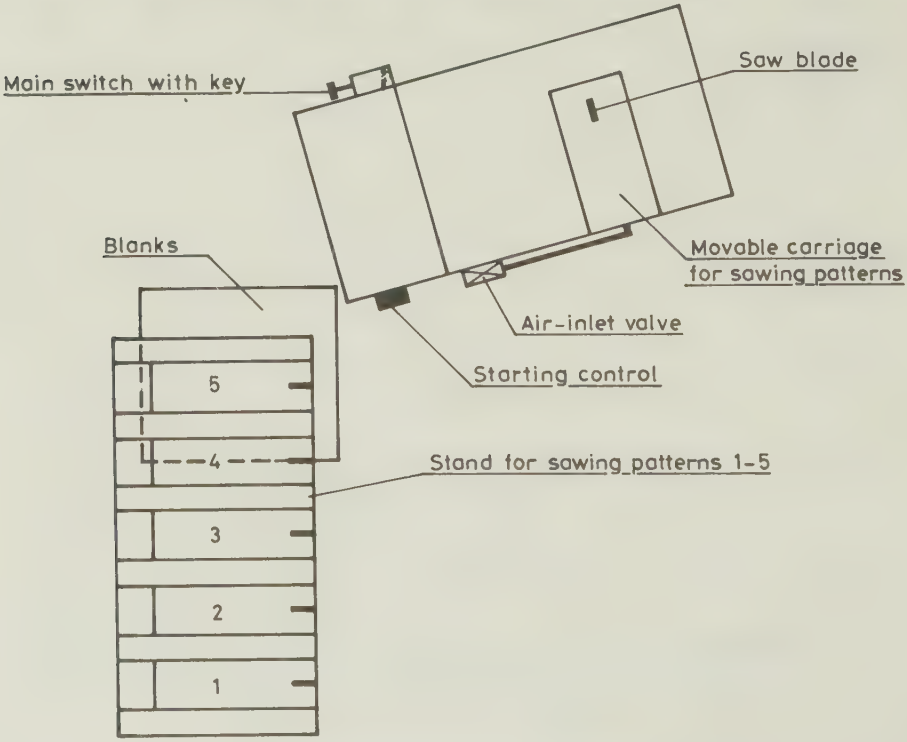



Fig. C 18. Sketch plan of workplace for sawing wooden frames.

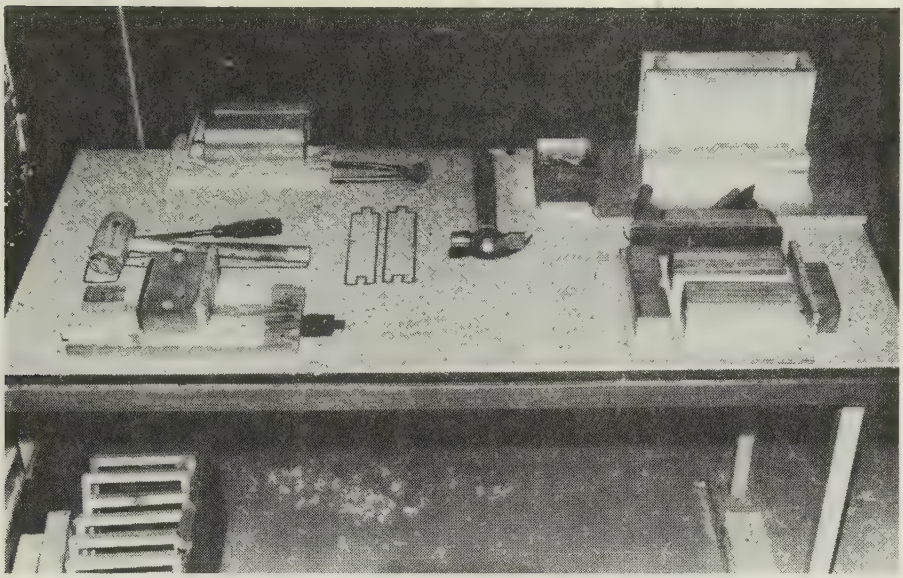


Fig. C 19. Photo of workplace for bench work on wooden frames.

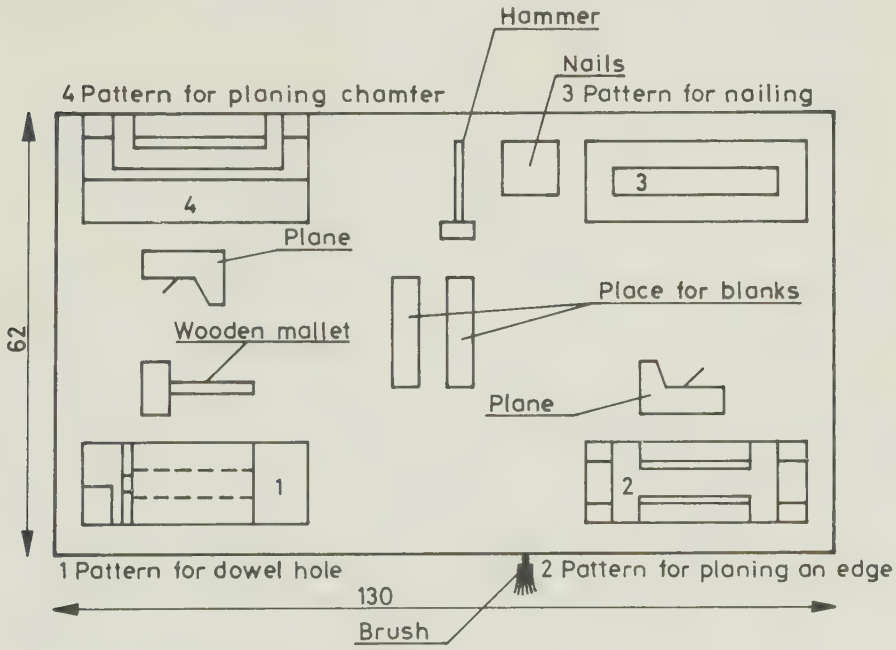



Fig. C 20. Sketch plan of workplace for bench work on wooden frames.

*Report on performances and development of performances.* In this task, the quantitative measurement of working capacity is calculated as the number of wooden frames produced and approved per hour during the last 2 working days. The distribution, mean value and variation of this measurement are shown in Table C6.

The distributions for both men and women display an obvious positive skewness, which indicates that the task is almost too difficult for these workers. This applies particularly to the women, who have a lower average level of performance than the men and who also present a smaller range.

The development of performances is shown in Fig. C16. This diagram confirms the above-mentioned sex difference in the level of performance. The men also seem to show a greater training effect. However, a clear training effect also appears among the women, and it is probable that the outputs would continue to increase for both men and women, if they were allowed to continue with the task for a longer time.

*Table C 6.* Number of wooden frames produced per hour.

| Details per hour | Men | Women |
|---|---|---|
| 0.0 – 0.4 | 8 | 12 |
| 0.5 – 0.9 | 7 | 6 |
| 1.0 – 1.4 | 4 | 2 |
| 1.5 – 1.9 | 7 | 3 |
| 2.0 – 2.4 | 1 | 1 |
| 2.5 – 2.9 | 4 | |
| 3.0 – 3.4 | 4 | |
| 3.5 – 3.9 | 3 | |
| 4.0 – 4.4 | 2 | |
| 4.5 – 4.9 | | |
| Σ | 40 | 24 |
| M | 1.7 | 0.7 |
| S.D. | 1.24 | 0.64 |

# 7. Embroidering cat motifs

*Description of workplace and cycle of operations.* The worker sits at a table, on one end of which a Husqvarna sewing machine (Automatic 21E7550) is placed. All the necessary working material is on the table. Cf. Figs. C22 and C23 on page 211.

The process begins with the worker taking from the plastic shelf on the table a piece of cloth (12 x 16 cm), cut and marked out in advance. An overcast seam is first sewed on it round the whole piece of cloth on the sewing machine. Then a cat motif is embroidered on it by hand. In order to facilitate the machine sewing, the maximum speed of the sewing machine is considerably reduced by reduction gearing and by fitting a locking device to the foot control. A special work fixture is available, in which the needle is placed to be threaded, which is done with the aid of a "needle threader". In front of the worker there are five patterns, which show different subsidiary stages in the embroidery work.

This task cannot be performed in a purely mechanical fashion. It includes several operations which require concentration and a certain understanding of the conditions of the process. Thus, in the machine sewing, the worker must concentrate on sewing along a straight line. He must not forget to move the sewn piece of cloth a little way from the presser foot before cutting the thread. In sewing by hand, it is a question of always aiming the needle with satisfactory precision, i.e. placing it within the small circular markings. It is also important that he should be able to pull the thread sufficiently tight.

According to the MTM expert, the task makes great demands on psychomotor ability and concentration. In other respects, it makes moderate demands on speed of movement, strength (in the fingers), persistence, ability to co-ordinate and ability to follow a pattern of movement. The training time is stated to be 2 weeks. After a further 2 weeks, the worker should have become used to the job. After so short a time as 4 days, the normal worker is expected to achieve 85 % of the MTM value. The making of one detail is calculated to take 6583 TMU or 3.95 minutes in the MTM analysis. The worker who achieves this rate will consequently produce 13.7 details per working hour.

*Evaluation of performances.* The production in each working period is noted and inspected. In this connection, the following procedure is applied.

The machined seam is first examined. For it to be considered satisfactory, it is necessary that the zigzag stitches shall always run outside the pencilled line. It is permissible for stitches to fall on the actual line but not for them to run inside it. The stitches must be sewn in the cloth and must accordingly not run out of it. However, the seam is approved if—as sometimes happens—there are signs that a thread has come undone from the cloth and has run together with the sewn stitches.

In examining the embroidery, the following criteria apply, as regards the eye loop. No dots must be passed over, nor may there be more stitches than there are dots. The stitches must be sewn in the dots. This means that the needle must be placed within or in the little circular line which marks the dot. The same applies to stitches in the whiskers and mouth.

The stitches in the whiskers, mouth and eye loop must not be tightened too much. This is checked as follows. When the cloth is stretched with both hands as far as the stitches allow—from one long side to the other—all the "bubbles" in the cloth must be smoothed out.

The nature of any faults which may occur and the number of approved products are stated in the record.

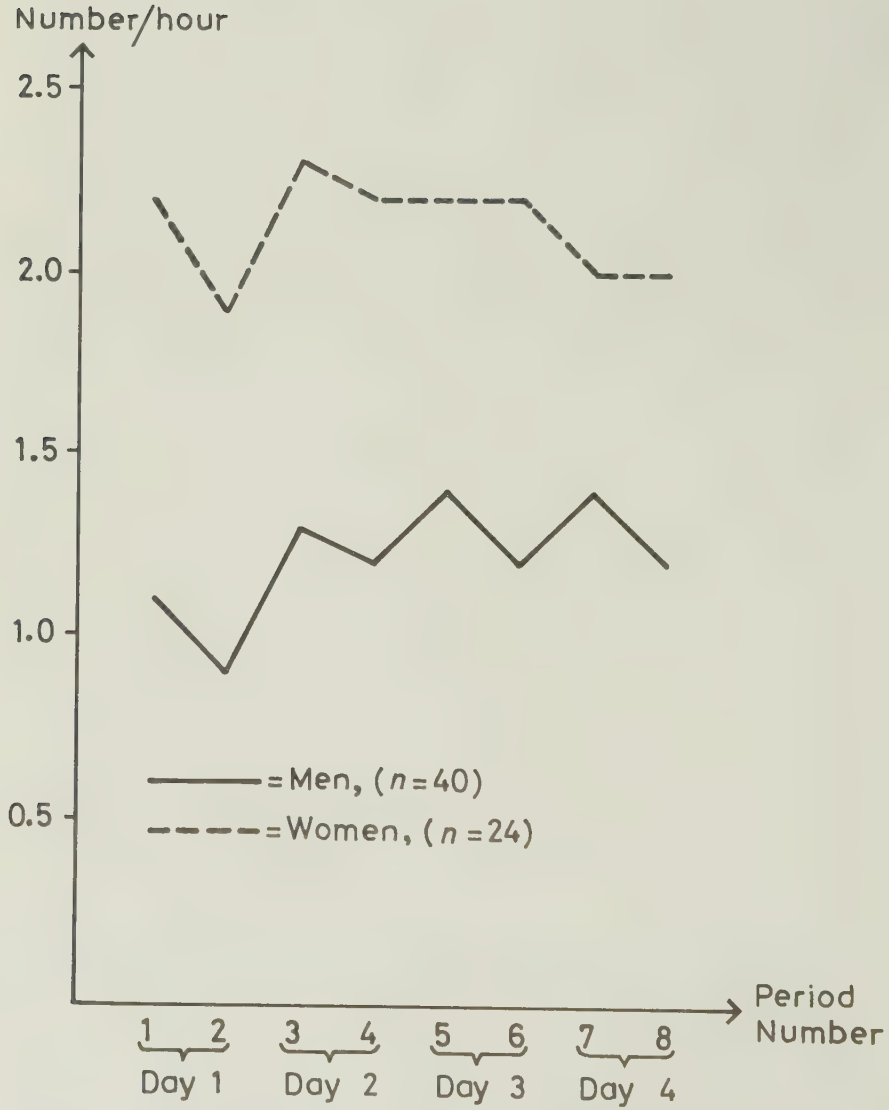



Fig. C 21. Development of performance in embroidering cat motifs.

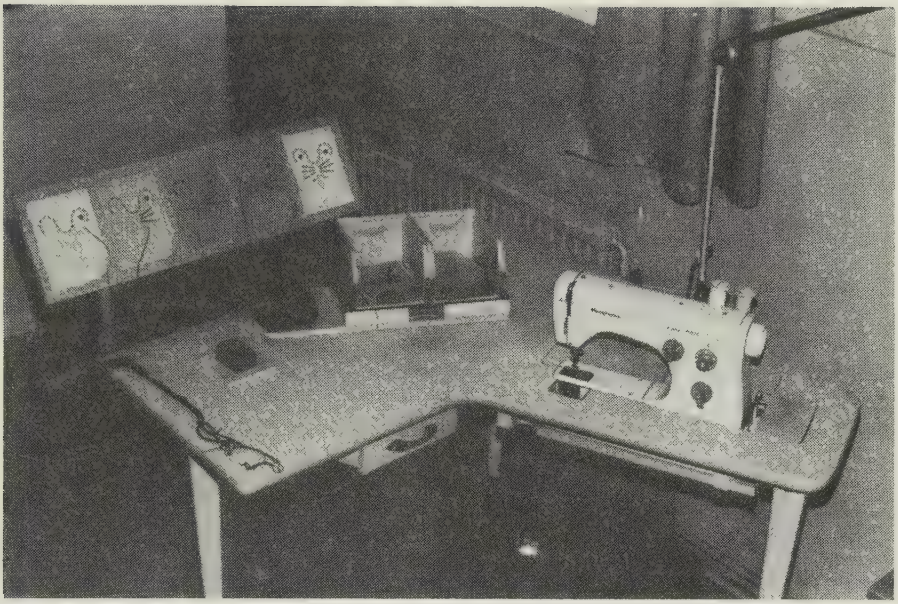


Fig. C 22. Photo of workplace for embroidering cat motifs.

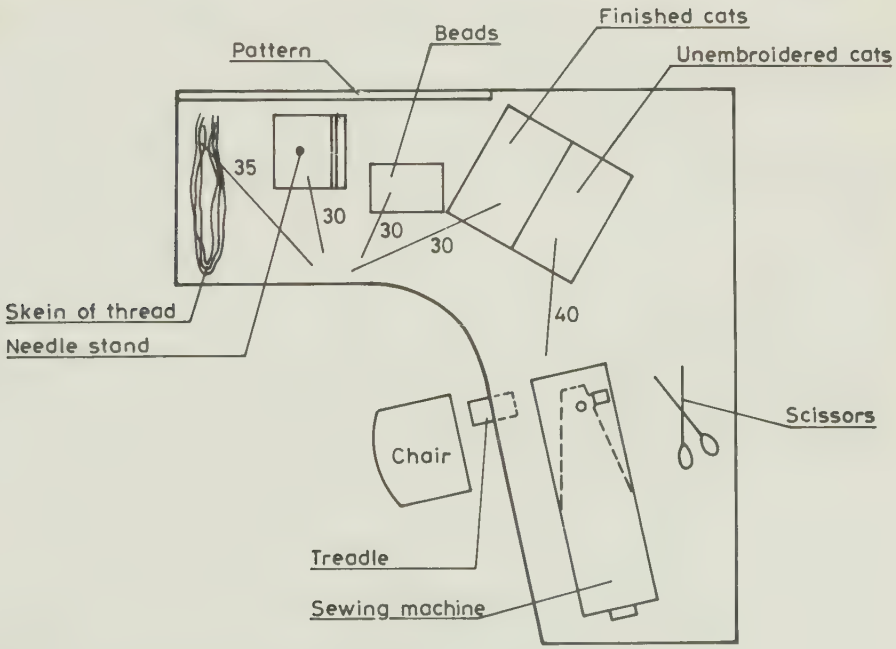



Fig. C 23. Sketch plan of workplace for embroidering cat motifs.

*Report on performances and development of performances.* By analogy with the procedure in the other tasks, the quantitative measurement of working capacity is here calculated as the number of approved details made per hour during the last 2 working days. The distribution, mean value and variation of this measurement are shown in Table C7.

As far as the men are concerned, the distribution is obviously skewed and indicates that the task gave them considerable difficulties. The women display a flatter distribution with a higher average value.

The development of performances during the 4 days is shown in Fig. C21. A common feature for both men and women is the decline in the outputs during the afternoon on the first day. This may be explained as a result of the considerable reduction of the intensive supervisory work done during the morning. The workers then had obvious difficulties in getting started on the task in a more independent manner. During the second day, they succeeded better in this and increased their outputs. However, later on, these outputs did not undergo any great change.

*Table C 7.* Number of approved cat motifs sewn per hour.

| Details per hour | Men | Women |
|---|---|---|
| 0.0 – 0.2 | 12 | 4 |
| 0.3 – 0.5 | 6 | 1 |
| 0.6 – 0.8 | 3 | 2 |
| 0.9 – 1.1 | 2 | 3 |
| 1.2 – 1.4 | 1 | |
| 1.5 – 1.7 | 1 | |
| 1.8 – 2.0 | 6 | 2 |
| 2.1 – 2.3 | 2 | |
| 2.4 – 2.6 | 1 | 2 |
| 2.7 – 2.9 | | 2 |
| 3.0 – 3.2 | 1 | 1 |
| 3.3 – 3.5 | 1 | 3 |
| 3.6 – 3.8 | 1 | 1 |
| 3.9 – 4.1 | | |
| 4.2 – 4.4 | 2 | 2 |
| 4.5 – 4.7 | | 1 |
| 4.8 – 5.0 | | |
| 5.1 – 5.3 | | |
| 5.4 – 5.6 | 1 | |
| Σ | 40 | 24 |
| M | 1.3 | 2.1 |
| S.D. | 1.41 | 1.48 |

# 8. Packing blocks

*Description of workplace and cycle of operations.* The worker has first to fold a carton and place it in a work fixture. Blocks are then packed in the carton in four different layers. Each layer has a special pattern, consisting of red and blue blocks of rectangular and triangular shapes. The triangular shape occurs only in two of the layers. A paper insert is placed between each layer. The four different layers are set up as a pattern on a board at eye level in front of the worker. The work is carried out sitting at a table, which measures 160 x 62 x 78 cm. In order to bring about a comfortable working position, the table top is tilted somewhat (about 10°), and in order to facilitate the work, there is a work fixture for the carton, as was mentioned above. The blocks are kept in four plastic boxes. Cf. Figs. C25 and C26 on pages 214—215.

In order to be able to perform this task in the approved manner, the worker must know all the time that he is laying the right layer and that the layer which he is laying in the carton is actually in accordance with the pattern. The packing of the triangularly shaped blocks seems to make special demands for a certain measure of spatial ability. For some of our workers, the task involved very great difficulties in this respect.

According to the MTM expert, the task makes very large demands for concentration, persistence and speed of movement. The demands for psychomotor ability and ability to co-ordinate are moderate, while the requirement as regards strength is evidently small. The training time for a normal worker is estimated to be a week. After a further week, the worker may be considered to be used to the job. In the MTM analysis, the packing of a carton is calculated to take 1894.1 TMU or 1.13 minutes. The worker who achieves this rate will consequently produce 47.5 details per working hour.

*Evaluation of performances.* The production in each working period is noted and all the cartons are inspected. Errors in the form of deviations from the prescribed patterns or the omission of the paper insert result in the cartons being rejected.

*Report on performances and development of performances.* The quantitative measurement of working capacity in this task is calculated as the number of correctly packed cartons per hour during the last 2 working days. The distribution, mean value and variation of this measurement are shown in Table C8.

The distributions show a certain tendency to positive skewness. The average level of performance is only slightly higher in men than in women. The ranges in both cases are large.

The development of performances during the 4 working days may be studied in Fig. C24. The curves reveal that training effects occur and seem to manifest themselves on approximately the same scale in both men and women. As I mentioned above, the average level of performance is also practically the same, but there is a difference in that the women seem to show a greater tendency to fluctuations from one working period to another.

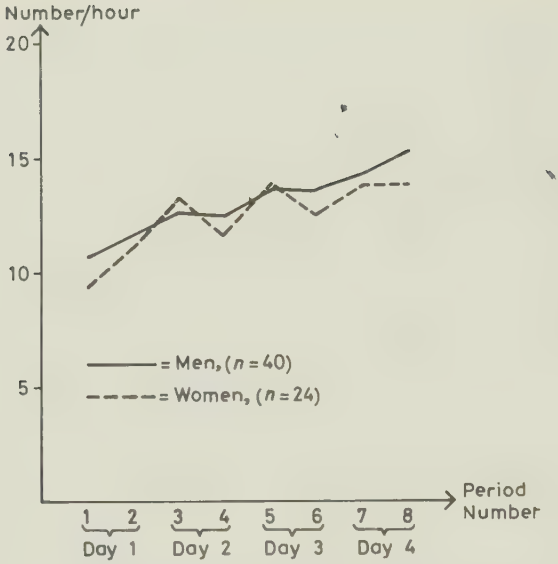



Fig. C 24. Development of performance in packing blocks.

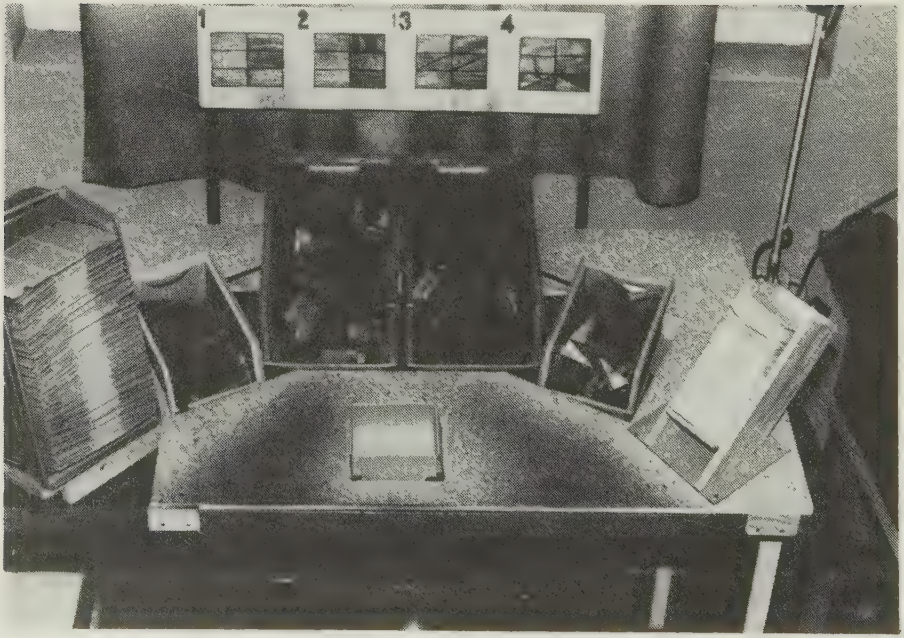



Fig. C 25. Photo of workplace for packing blocks.

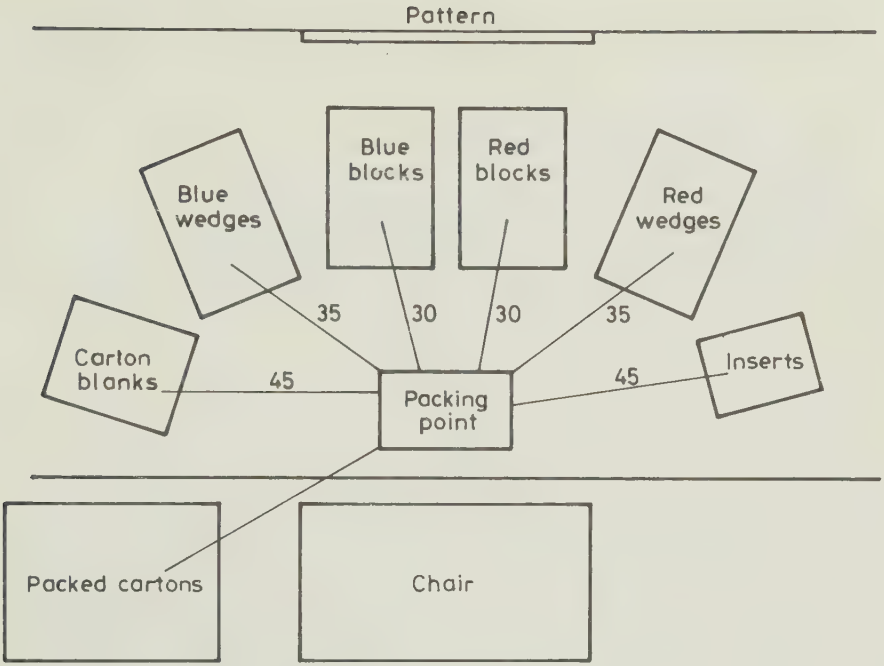



Fig. C 26. Sketch plan of workplace for packing blocks.

*Table C 8.* Number of cartons packed per hour.

| Details per hour | Men | Women |
|---|---|---|
| 0 – 2 | 4 | 1 |
| 3 – 5 | 3 | 5 |
| 6 – 8 | 7 | 1 |
| 9 – 11 | 3 | 4 |
| 12 – 14 | 4 | 2 |
| 15 – 17 | 6 | 4 |
| 18 – 20 | 4 | 2 |
| 21 – 23 | 2 | 1 |
| 24 – 26 | 3 | 2 |
| 27 – 29 | 3 | 1 |
| 30 – 32 | | |
| 33 – 35 | | 1 |
| 36 – 38 | | |
| 39 – 41 | 1 | |
| Σ | 40 | 24 |
| M | 14.2 | 13.6 |
| S.D. | 8.98 | 8.55 |

# Appendix D.

## Control Study of the Extent of Possible Transfer Effects on the Measurement of Performance

The background of this control study is reported in Chapter IV on p. 000, where there is a discussion of the appropriate sequence of the eight tasks included in the training and testing programme. In deciding whether to present these tasks in one and the same order to all experimental subjects or to apply a rotation system, which meant that each task was presented just as often as nos. 1, 2, 3, etc., I chose the latter alternative for various reasons. However, I was aware that the application of such a rotation system might involve a disadvantage, insofar as transfer effects occurred. Such effects might lead to unfair comparisons between individual performances. The quantitative measurements of performance serving as criteria would consequently be described as possessing the kind of heterogeneity against which Thorndike (1963) warned us.

Guided by this insight, I tried to reduce the scope of possible transfer effects by staggering the training and testing periods. The experimental subject had to return to his or her ordinary job for a week between two such periods. At the same time, however, I judged it desirable to try to obtain a measurement of the magnitude of the transfer effects which might conceivably arise despite this arrangement. For this purpose, the eight tasks were divided into comparable groups of four. In the same way, the experimental subjects were divided into two groups of equal size which were made as similar as possible in a number of respects by matching as regards sex, length of preliminary training, estimated previous experience and performance in tests of fine motor function. During the first half of the training and testing programme, one of these experimental groups was assigned to work on cutting threads in nuts, turning holders, assembling electrical panels and packing blocks. These experimental subjects where then allowed, in the second half of the programme, to transfer to pressing eyelets, making wooden frames, embroidering cat motifs and assembling brackets. The second group of experimental subjects went in the opposite order and thus began by completing the tasks of pressing eyelets, making wooden frames, embroidering cat motifs and assembling brackets, before they began the second half of their training and testing programme by changing to threading nuts, turning holders, assembling electrical panels and packing blocks. The entire scheme is illustrated by the sketch in Fig. 9 on p. 61.

If transfer effects did occur, this should be evident as a difference in performance between the two groups of experimental subjects. Thus, Group 1 should show a somewhat lower performance in the first group of tasks and a somewhat higher one in the second—in comparison with Group 2.

Before reporting the results of the control study, I shall comment on the procedure for matching the two experimental groups. As the description of the implementation of the testing programme suggests, four groups of 16 workers each were chosen to carry out the programme in turn. For various reasons, it was desirable to match these groups against each other. Once this was done, half the 16 workers in the respective groups were assigned—with due consideration of the individual values in the matching variables—to Group 1 in the control study and the other half to Group 2. In this connection it did not prove to be possible, with a reasonable amount of success, to make the two groups equal as regarded length of training, so that this matching variable had to be ignored. The result of the matching was otherwise satisfactory—as Table D 1 shows—althought with the exception that the standard deviations are somewhat lower in Group 2.

Let us now, with the aid of Table D 2, examine the results of the completed control study and assess whether or not transfer effects may have occurred in connection with the measurements of the quantitative performances in the tasks. This involves studying the divergences between the two groups of experimental subjects. As the next to the last column of the table shows, an analysis of significance was carried out. A correction was

*Table D 1.* Groups in the control study and their values in the matching variables.

| Experimental group | Sex | Prognosis group | | Percentage group | | Psychomotor ability | |
|---|---|---|---|---|---|---|---|
| | | M | S.D. | M | S.D. | M | S.D. |
| Group 1<br>$n = 32$ | Men, $n = 20$<br>Women, $n = 12$ | 3.88 | 1.05 | 2.53 | 1.25 | 15.78 | 6.42 |
| Group 2<br>$n = 32$ | Men, $n = 20$<br>Women, $n = 12$ | 3.94 | 0.79 | 2.63 | 1.19 | 15.47 | 5.20 |

*Table D 2.* Measurements of quantitative performances in two groups of experimental subjects for control of possible transfer effects and their magnitude.

| | Gropups of tasks | Groups of subjects | | | | *t* values for obtained differences | Expected differences if positive transfer effects occur |
|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | | |
| | | M | S.D. | M | S.D. | | |
| I | Threading nuts | 186.22 | 72.74 | 190.50 | 63.46 | 0.37 | Lower values in group I |
| I | Turning holders | 41.41 | 21.97 | 35.31 | 19.36 | 1.66 | |
| I | Assembly of electrical panels | 2.81 | 1.86 | 2.43 | 1.53 | 1.53 | |
| I | Packing blocks | 14.13 | 8.73 | 13.78 | 8.92 | 0.16 | |
| II | Pressing eyelets | 274.41 | 130.84 | 303.44 | 120.10 | 1.48 | Higher values in group I |
| II | Making wooden frames | 1.56 | 1.25 | 1.15 | 1.01 | 2.48 | |
| II | Embroidering cat motifs | 1.47 | 1.42 | 1.79 | 1.57 | 1.29 | |
| II | Assembly of brackets | 113.78 | 45.45 | 119.31 | 46.02 | 0.62 | |

made for the multiple correlation between the four matching variables and the test variable, as recommended by Garrett (1947, p. 213). I may first state that only in two cases are the differences between the two groups of experimental subjects in the direction to be expected if positive transfer effects influenced the results. These concerned cutting threads in nuts, in which the difference is negligible, and making wooden frames. In the latter task, the difference is striking and reaches statistical significance at the 2 % level. In the other six tasks, the divergences are in the opposite direction, so that there seems reason to conclude that positive transfer effects only occurred in making wooden frames. Such a result is not in itself wholly unexpected. Making wooden frames is the task calling for the longest time and containing the largest number of stages, which are furthermore greatly varied. Thus, the possibilities of benefiting by experience from other tasks may be especially good.

The six differences, which have an opposite trend to that which may be created by positive transfer effects, deserve special comment. None of them reach statistical significance, and one may therefore be justified in denying them any real importance. However, the fact that the differences run parallel in no less than six cases out of eight may give grounds for the hypothesis that the performances during the second part of the training and testing programme were adversely influenced by fatigue or declining motivation on the part of the experimental subjects. Such a hypothesis would agree with the conclusion drawn from the special studies reported in another appendix, concerning the effects of training for longer periods, namely, that performances among mentally retarded seem, in the long run, to be influenced by motivation factors rather than by training factors. However, even on this point there is reason to qualify such an interpretation with definite reservations and to strongly emphasize that a more definitive illustration of this vital problem of the development of performance among the mentally retarded during longer periods can only be obtained by large-scale new empirical investigations.

Thus, the chief conclusion from the control study reported here is that positive transfer effects only influenced the results of the criterion measurements to a very limited extent in these enquiries. Signs of such influence only appeared in making wooden frames. However, one cannot entirely ignore the possibility that certain negative effects, caused by incipient fatigue or waning motivation may, on the other hand, have resulted in a decrease of the performances in some of the other tasks. Yet, it is in any case a question of minor effects, which we may well assume did not contribute to any crucial extent to reduce the homogeneity of criteria or to make individual comparisons impossible.

# Appendix E

## Effects of Training over a Long Period

In this investigation of performance and adjustment to work among mentally retarded adults, I measured quantitative performances in eight selected tasks, by calculating the number of approved products per hour during the last two of four working days. Apart from reaching high standards in terms of reliability, these measurements thus have the virtue of expressing the performances which the workers could achieve after receiving thorough instruction and training in the respective tasks for two whole days. However, as regards the latter viewpoint, there are good grounds for objecting that the extent of this training was insufficient and that there was some likelihood that the workers could, after further training, improve their performance, which would also alter their ranking order.

Experience often confirms that every individual needs practice—and that to an extent related to the nature of the tasks—before reaching the normal performance given by the MTM value. In Appendix C, I have reported the assessments which the technical expert made of the probable appearance of the courses of practice among normal workers on the eight tasks. I can here sum up by saying that from this point of view these tasks can be divided into two groups. Cutting threads in nuts, turning holders, assembling brackets, and packing blocks are all regarded as needing only short periods of practice, so that one may assume that normal performance will be achieved after one week's work or less. The other four tasks are judged to demand more practice, and after a period as short as four days one can expect that the normal worker will only achieve a certain percentage of the MTM values. This figure was set at 90 % for pressing eyelets, 60 % for assembling electrical panels, 75 % for making wooden frames and 85 % for embroidering cat motifs.

Notwithstanding that these asessments were made from considerable knowledge and that we can be fairly certain that, on the whole, they give an accurate picture of reality, it is, of course, essential to carry out empirical studies, if their accuracy is to be scientifically verified.

The situation is no less, but rather more, complicated in the case of mentally retarded workers. One cannot directly apply to such workers assessments of the kind described above, and in fact it is very frequently assumed that in their case the practice periods must be much longer before maximum performance is attained. I found it the more important to try, at least to some degree, to give an empirical illustration of the problems suggested here. For this reason, I made two studies, the extent of which had to be strictly limited in relation to my original plans, because the workshop in which I carried out my studies was closed down.

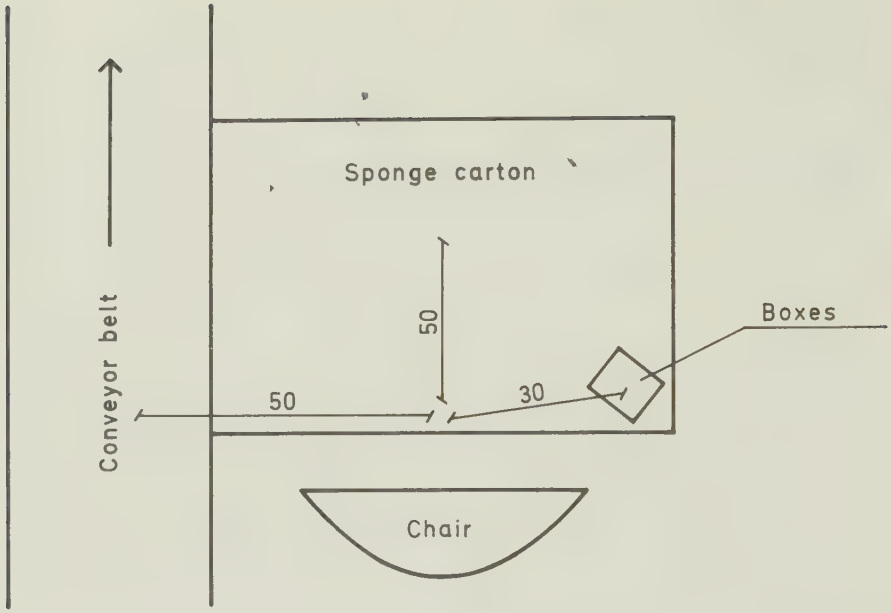



Fig. E 1. Sketch plan of workplace in the sponge investigation. Distances of motion given in centimetres.

## The sponge investigation

In the first of these studies, I selected a task long included in the production of the ALA Foundation workshop, namely packing of household sponges. This may be described as an undemanding task with a short cycle; it is carried on at a number of stations along a conveyor belt (cf. Fig. E1 above). The workers had first to pick up a box from the pile, open it, and fold back two flaps. Then he took a sponge from the carton in front of him, put it into the box, and finally placed the packed sponge on the conveyor belt. My technical colleague carried out an MTM analysis of the task and we also made special arrangements to measure the performances in accordance with the same principles as applied in the main investigation. Twelve men and 13 women workers, all of whom had previously packed sponges for at least three months, were chosen to take part in the experiment. Thus, they had had extensive training and were used to this task.

These participating in the experiment were prepared and stimulated in the usual way to do their best. Their performances were then measured for two consecutive days and expressed as the number of sponges packed per hour. Table E1 gives the distribution of the results for both men and women.

*Table E 1.* Number of sponges packed per hour.

| Details per hour | Men ($n$ = 12) | Women ($n$ = 13) |
|---|---|---|
| 116 – 135 | 2 | |
| 136 – 155 | 1 | 1 |
| 156 – 175 | | 1 |
| 176 – 195 | 2 | 2 |
| 196 – 215 | 1 | 1 |
| 216 – 235 | | |
| 236 – 255 | 1 | 2 |
| 256 – 275 | 1 | 2 |
| 276 – 295 | 2 | 2 |
| 296 – 315 | 2 | |
| 316 – 335 | 1 | |
| 336 – 355 | | 1 |
| 356 – 375 | | |
| 376 – 395 | | |
| 396 – 415 | | |
| 416 – 435 | | |
| 436 – 455 | | |
| 456 – 475 | | 1 |
| | M 230 | 254 |

According to the completed MTM analysis, a normal worker with sufficient practice could pack 606 sponges per hour. This means that, on the average, the male workers achieved a standard equivalent to 38 % of the normal performance, while the female workers achieved 42 % thereof.

In order to interpret this result, I shall compare it with a corresponding assessment of the performances in the eight tasks included in the main investigation. In Table E2 I have expressed these as percentage work contributions in relation to the normal performances expected according to calculated MTM values.

In the table the tasks are divided into two groups. One includes cutting threads in nuts, turning holders, assembling brackets and packing blocks, or those tasks which were assessed by the MTM expert as demanding least practice. The other group includes pressing eyelets, assembling electrical panels, making wooden frames and embroidering cat motifs, which were thought to need two weeks or more of practice before the MTM value was reached by the normal worker.

The figures in the table show that the percentage work contribution in the least demanding tasks averages 43 % (men) and 35 % (women) respectively, while in the more demanding tasks it reaches 27 % and 20 % respectively. These figures suggest that the assessment by an MTM expert of the training necessary for the different tasks also has a certain validity as far as the mentally retarded are concerned. The essential question now, however, is how great an effect on these performances would be achieved by further training of the workers.

This question may be illustrated to some extent by the investigation presented above, to which I shall now return. I have already pointed out that packing sponges is a short-cycle, undemanding task. From the point of view of training it is comparable with cutting threads in nuts, turning holders, assembling brackets and packing blocks. The percen-

*Table E 2.* Percentage work contributions in different tasks.

| Tasks | Men | | Women | |
|---|---|---|---|---|
| | Total $n = 40$ | Participants in "sponge investigation" $n = 12$ | Total $n = 24$ | Participants in "sponge investigation" $n = 13$ |
| 1. Threading nuts | 61 | 54 | 52 | 54 |
| 2. Turning holders | 38 | 29 | 24 | 22 |
| 3. Assembling brackets | 42 | 38 | 36 | 35 |
| 4. Packing blocks | 30 | 21 | 29 | 24 |
| $M_{1-4}$ | 43 | 36 | 35 | 34 |
| 5. Pressing eyelets | 35 | 28 | 35 | 33 |
| 6. Assembling electrical panels | 25 | 18 | 16 | 15 |
| 7. Making wooden frames | 39 | 23 | 17 | 14 |
| 8. Embroidering cat motifs | 10 | 5 | 15 | 13 |
| $M_{5-8}$ | 27 | 19 | 20 | 19 |

tages for the group of workers included in the "sponge investigation" are given in a special column in Table E2. We find that their average percentual effort in these less demanding tasks amounts to 36 % (men) and 34 % (women) respectively. The corresponding figures for packing sponges were 38 % and 42 % respectively. This would suggest that a longer period of training in such tasks may lead to a certain improvement in performance but that it seems nevertheless to be of comparatively limited extent.

## Development of performance in assembling electrical panels for altogether four weeks

The conclusion drawn above, however, is in fact rather uncertain, partly because it was not possible in this study to make comparisons between performances after shorter or longer periods of practice at one and same task. In the second of the two investigations carried out to illustrate the question of long-term effects, I am in a better position in this respect. I allowed 18 male and eight female workers to spend their time assembling electrical panels for three consecutive weeks, during which I studied the development of their performances in the same way as in the previous main investigation. Since all the subjects of the experiment also took part in the main investigation, this means that they received altogether four weeks of training in this task, which, according to the MTM analytical assessement, is precisely the period for a normal worker to reach a standard equal to the norm. Table E3 gives the calculated quantitative measurements of performance during the four working weeks. The development of performance during the same period can be followed in more detail in Fig. E2. It should be remembered that, between the first working week and the following three-week period, there was an interval varying in length from four to 18 months for the different subjects.

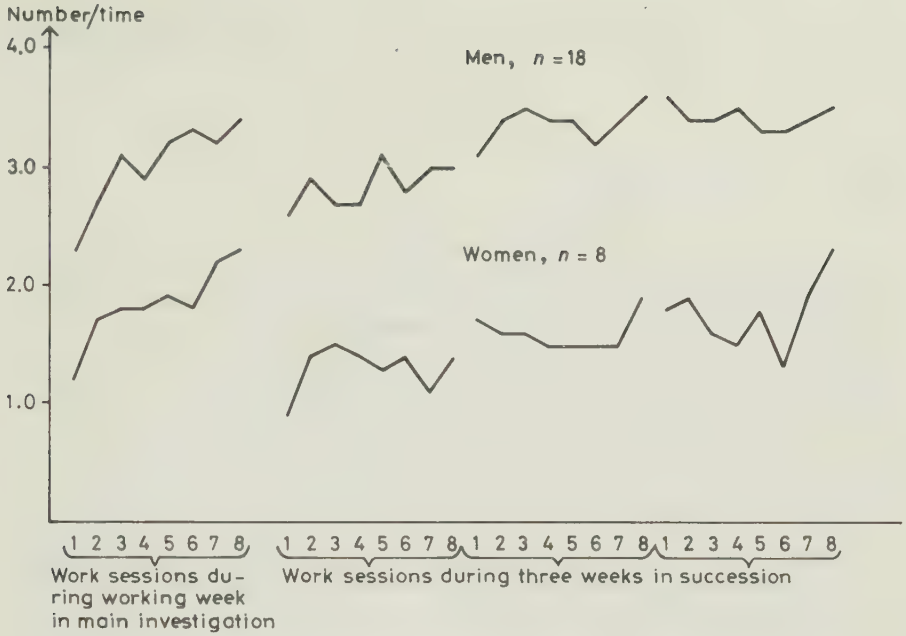

Number/time
4.0
3.0
2.0
1.0
Men, n = 18
Women, n = 8
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
Work sessions during working week in main investigation
Work sessions during three weeks in succession

Fig. E 2. Development of perfomance among men (upper line) and women (lower line) in assembly of electrical panels during altogether four working weeks.

*Table E 3.* Average quantitative performances in assembly of electrical panels during four weeks in men and women.

| Week | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Men ($n$ = 18) | 3.3 | 3.0 | 3.4 | 3.4 |
| Women ($n$ = 8) | 2.0 | 1.3 | 1.6 | 1.8 |

It is characteristic of both men and women that the performances at the beginning of the three-week period are lower than those in the main investigation. Afterwards a slight rise may be noted, which appears to be consistent and convincing in the male group but in the female group is slower and less regular in tempo than during the first working week. Towards the end of the three-week period, the men exceeded the standard which they reached in the main investigation, while the women fell short of it. If the performance level is expressed as the average percentage work contribution (in relation to the MTM norm) and this is calculated on the basis of the last two days in the week concerned, it amounts during the first week to 26 % (men) and 16 % (women), while the corresponding figures for the fourth and last week are 29 % and 15 % respectively. This means that a limited effect of training can be discerned in the men, while in the women this effect is conspicuous by its absence. The calculated product-moment correlation between the quantitative performances during the first and last weeks amounts to .77,

which means that certain individual shifts occurred but that there was nevertheless comparatively close agreement between the ranking orders on the two occasions when measurements were made.

## Discussion and conclusions

The study here reported of the development of performance during a long period of work on one and the same task yielded results pointing in the same direction as the sponge investigation. Judging from both studies, it would seem that practice on this type of task does little to improve performance once the initial difficulties are overcome. These results may seem surprising in view of the often-stated belief that mentally retarded people can considerably improve their performance by training, if only this training is of sufficient extent. I do not claim that my investigations have refuted this view in the general sense. The problem deserves a much more all-round and thorough empirical illustration. In this respect it is especially necessary to pay great attention to the part which motivation to work plays in performance. My systematic observations of behaviour indicated that the motivation to work in the assembly of electrical panels was at a very high level during the main investigation but showed a tendency to wane during the three-week period studied. It is not out of the question that we could have noted somewhat higher performances towards the end of that period if the motivation could have been maintained at the same high level throughout the course of the work. On the other hand, there is good reason to ask which motivation level is realistic or rooted in reality, in the sense that it can be maintained without artificial means for a long period of continuous work in a normal environment.

In a special investigation I have discussed in detail the question of the correlation between motivation and performance (Holm and Palmér 1972). In the present context I would sum up by saying that the results reported above support my view that the quantitative measurements of performance which I calculated in the main investigation are to be regarded as realistic expressions of the standard which the subjects may be expected to attain after basic training in a favourable environment. In the longer term, variations in performance would seem to be determined more by motivation factors rather than by training factors. This does not, of course, exclude the possibility that further training would probably result in improved performances by certain individuals, especially in tasks with a longer cycle, such as making wooden frames. However, the changes in the ranking orders of the individual workers would remain within fairly narrow limits.

# References


Baumeister, A.A. & Kellas, G. Reaction time and mental retardation. In Ellis, N.R. (ed.) *International Review of Research in Mental Retardation,* Vol. 3, 1968.

Bemis, S.E. Occupational validity of the general aptitude test battery. *J. Appl. Psychol., 52*:240–244, 1969.

Benton, A.L. Psychological evaluation and differential diagnosis. In Stevens, H.A. & Heber, R. (eds.) *Mental Retardation,* The University of Chicago Press, 1964.

Bodin, I. & Willquist, I. *Psykiskt utvecklingsstörda ungdomars arbetsförmåga och arbetsanpassning.* Psykologexamensarbete vid Psykologiska institutionen, Uppsala universitet, Uppsala, 1972.

Bray, D.W. & Moses, J.L. Personnel selection. In *Ann. Rev. Psychol. 23*:545–576, 1972.

Burket, G.R. A study of reduced rank models for multiple prediction. *Psychometric Monograph,* no. *12,* 1964.

Bäckström, L. Elsa. A Fortran program for statistical analysis. *Report 78 from Department of Psychology, University of Uppsala,* Sweden (mimeographed), 1970.

Clarke, A.M. & Clarke, D.B. The abilities and trainability of imbeciles. In Clarke, A.M. & Clarke, D.B. (eds.), *Mental Deficiency. The Changing Outlook,* (2nd ed.), New York: The free press, 1965.

Cobb, H.V. *The Predictive Assessment of the Adult Retarded for Social and Vocational Adjustment.* A Review of Research, Part I, Annotated Bibliography (mimeographed), Department of Psychology, The University of South Dakota, 1966.

– *The Predictive Assessment of the Adult Retarded for Social and Vocational Adjustment.* A Review of Research, Part II, Analysis of the Literature, 1969 (mimeographed, also published in Teachers College Press, 1972).

Cramer, E.M. Significance tests and tests of models in multiple regression. *The American Statistician, 26:* 26–30, 1972.

Cronbach, L.J. *Essentials of Psychological Testing,* (2nd ed.), Harper and Row, New York, 1969.

– Test validation. In Thorndike, R.L. (ed.), *Educational Measurement* (2nd ed.), American Council on Education, Washington, D.C., 1971.

Dingman, H.F. & Meyers, C.E. The structure of intellect in the mental retardate. In Ellis, N.R. (ed.) *International Review of Research in Mental Retardation,* Vol. 1, 1966.

Domme, Ch. & Lundahl, B. *Konstruktion och preliminär utprövning av prov i matematik för särskolan.* Pedagogiska institutionen, Uppsala (mimeographed), 1966.

Draper, N.R. & Smith, H. *Applied Regression Analysis,* John Wiley & Sons, Inc., 1966.

Dureman, I. & Sälde, H. *Experimentalpsykologiska metoder för klinisk tillämpning,* Acta Acad. Reg. Sci., 2, Almqvist & Wiksell, Uppsala, 1959.

Ebel, R.L. Estimation of the reliability of ratings, *Psychometrika, 16:* 407–424, 1951.

Edwards, A.L. *Experimental Design in Psychological Research* (3rd ed.), New York, 1968.

Elfström, M. & Granberg, M. *Träning av motoriska färdigheter hos mentalt retarderade och normalbegåvade.* Rapport från Psykologavdelningen, Ulleråkers sjukhus, Uppsala (mimeographed), 1968.

Elkin, L. Predicting productivity of trainable retardates on experimental workshop tasks. *Amer. J. Ment. Defic. 71:* 576–580, 1967.
– Predicting performance of the mentally retarded on sheltered workshop and non-institutional jobs. *Amer. J. Ment. Defic., 72:* 533–539, 1968.
Fleishman, E.H. & Hempel, W.E. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika, 19*:239–252, 1954.
Fränkel, F., Hedsjö, C. & Palmér, R. Rekrytering av arbetstagare till skyddad verkstad för psykiskt utvecklingshämmade. *Supplementum nr 4 till Psykisk Utvecklingshämning,* p. 213, 1969.
Garrett, H.E. *Statistics in Psychology and Education* (3rd ed.), Longmans, Green and Co, 1947.
– A developmental theory of intelligence. *Amer. Psychologist, 1:* 372–278, 1946.
Geary, R.C. & Leser, C.E.V. Significance tests in multiple regression. *The American Statistician, 22:* 20–21, 1968.
Goddard, H.H. Suggestions for a prognostic classification of mental defectives. *J. Psycho-Asthenics, 14:* 48–54, 1909.
Gold, M.W. Research on the vocational habilitation of the retarded: The present, the future. In Ellis, N.R. (ed.) *International Review of Research in Mental Retardation,* Vol. 6, 1973.
Goldstein, H. Social and occupational adjustment. In Stevens, H.A. & Heber, R. (eds.) *Mental Retardation,* The University of Chicago Press, 1964.
Grossman, H.J. (ed.) Manual on terminology and classification in mental retardation, 1973 revision, *American Association on Mental Deficiency. Special Publication Series No. 2,* 1973.
Grunewald, K. *The Mentally Retarded in Sweden.* Published by the Swedish Institute, 1969.
Guilford, J.P. *Psychometric Methods* (2nd ed.), McGraw-Hill Book Company, New York, 1954.
– *Fundamental Statistics in Psychology and Education* (4th ed.), McGraw Hill Book Company, New York, 1965.
Guskin, S.L. & Spicker, H.H. Educational research in mental retardation. In Ellis, N.R. (ed.) *International Review of Research in Mental Retardation,* Vol. 3, 1968.
Hasselqvist, O., Söderström, P. & Wiklund, A. *MTM:s grundrörelser* (3rd ed.), Svenska MTM-föreningen AB, 1967.
Hector, K.G. & Holmkvist, N.O. *Manual till Motor Test of Tracing Ability* (MTTA), Skand. Testförlaget AB, Stockholm, 1966.
Henrysson, S. *Applicability of Factor Analysis in the Behavioral Sciences.* Almqvist & Wiksell, Stockholm, 1957.
Holm, G. & Palmér, R. *Arbetsmotivation och arbetsprestation hos mentalt retarderade.* Rapport från Pedagogiska institutionen vid Lärarhögskolan i Uppsala (mimeographed), 1972.
Holmlund, M., Johansson, L., Nordin, K. & Ryberg, K. *Undersökning av journal- och testdatas prognosvärde för utskrivning av mentalt retarderade kvinnor.* Rapport från Psykologavdelningen, Ulleråkers sjukhus, Uppsala, 1971.
Horst, P. *Psychological Measurement and Prediction.* Wadsworth Publishing Company, Inc. 1966.
Jungner, R. Särskolan (sinnesslöskolan). In Norman, H. (ed.) *Mentalhygien och pedagogik,* Bokförlaget Natur och Kultur, Stockholm, 1955.

Jöreskog, K.G. *Statistical Estimation in Factor Analysis,* Almqvist & Wiksell, Uppsala, 1963.

Kebbon, L. *The Structure of Abilities at Lower Levels of Intelligence. A Factor-Analytical Study,* Skand. Testförlaget, Stockholm, 1965.

— *US-Batteriet. Manual.* Skand. Testförlaget, Stockholm, 1968.

— Vilka skrivs ut från sjukhus för utvecklingshämmade? *Psykisk Utvecklingshämning, 4,* 1969.

Kebbon, L., Aldrin, M.E. & Hallerstedt, G. The structure of abilities in the retarded. *Scand. J. Psychol., Vol. 8:* 150–154, 1967.

Kokaska, C.J. The occupational status of the educable mentally retarded: A review of follow-up studies. *The Journal of Special Education, vol. 2/no. 4,* 1968.

Kääriäinen, R. The factor structure of intellectual abilities and signal sight vocabulary learning at moderate and severe levels of preliterate mental retardation. *Research Bulletin, no. 3, Department of Educational Research, Gothenburg School of Education,* 1970.

Lindgren, A. Mental and social development of patients with Down's syndrome. In *Proceedings of the Second Congress of the International Association for the Scientific Study of Mental Deficiency,* 1971.

Lord, F.M. & Novick, M.R. *Statistical Theories of Mental Test Scores,* Reading, Mass., Addison-Wesley Publishing Company, 1968.

McCarron, L.T. & Dial, J.G. Neuropsychological predictors of sheltered workshop performance. *Amer. J. Ment. Defic., 77:* 241–243, 1972.

McNemar, Q. *Psychological Statistics* (4th ed.), John Wiley and Sons, Inc. 1969.

Mischel, W. *Personality and Assessment,* New York: Wiley, 1968.

Nilsson, I. & Nordlund, G. Prognos av framgång i gymnasiet IV. En studie med differentierade prediktorer och kriterier. *Pedagogiska rapporter, Umeå, 26,* 1972.

Nordgren, B. Physical capabilities in a group of mentally retarded adults. *Scand. J. Rehab. Med., 2:* 125–132, 1970.

Nordgren, B. & Bäckström, L. Correlations between muscular strength and industrial work performance in mentally retarded persons. *Acta Paediat. Scand. Suppl. 217,* 1971.

Nordström, S.G. *Hjälpskolan och särskolan i Sverige t.o.m. 1921,* Årsböcker i svensk undervisningshistoria, 119, 1968.

Nunnally, J.C. *Psychometric Theory,* McGraw-Hill Book Company, New York, 1967.

O'Connor, N. & Tizard, J. Predicting the occupational adequacy of certified mental defectives. *Occup. Psychol., 25:* 205–211, 1951.

— *The Social Problem of Mental Deficiency,* London: Pergamon Press, 1956.

Palmér, R. Skyddade verkstäder för psykiskt utvecklingshämmade. *Supplementum nr 2 för Psykisk Utvecklingshämning,* 1968 (*a*).

— Utgångspunkter och slutsatser. In Talsvårigheter och talträning hos psykiskt utvecklingshämmade. Ett symposium. *Supplementum nr 3 för Psykisk Utvecklingshämning,* 1968 (*b*).

— Arbetsprövning av utvecklingshämmade. *Psykisk Utvecklingshämning, 2,* 1971.

Palmér, R., Holm, G., Lööw, J.-O. & Ohlsson, C.-E. *Ett tränings- och prövningsprogram för psykiskt utvecklingsstörda arbetstagare.* Rapport nr 31 från forskningsprojektet Psykiskt utvecklingshämmade ungdomars yrkes- och samhällsanpassning (mimeographed), 1970.

Patterson, C.H. Methods of assessing the vocational adjustment potential of the handicapped. *The Training School Bulletin, 61:* 129–152, 1964.

Ronan, W.W. & Prien, E.P. *Toward a Criterion Theory. A Review and Analysis of Research and Opinion,* Greensboro, N.C.: Richardson Found, 1966.
Rosen, M., Kivitz, M.S., Clark, G.R. & Floor, L. Prediction of postinstitutional adjustment of mentally retarded adults. *Amer. J. Ment. Defic., 74:* 726–734, 1970.
Rosen, M., Floor, L. & Baxter, D. Prediction of community adjustment: A failure at cross-validation. *Amer. J. Ment. Defic., 77:*111–112, 1972.
Rubenowitz, S. *Personaladministration och arbetspsykologi* (2nd ed.), Akademiförlaget 1971.
Ryan, J. Classification and behaviour in mental subnormality–some implications for research. In *Proceedings of the Second Congress of the International Association for the Scientific Study of Mental Deficiency,* 1971.
Shafter, A.J. Criteria for selecting institutionalized mental defectives for vocational placement. *Amer. J. Ment. Defic. 61:* 599–616, 1957.
Siegel, P.S. Incentive motivation in the mental retardate. In Ellis, N.R. (ed.) *International Review of Research in Mental Retardation.* Vol. 3, 1968.
Stevens, H.A. Overview. In Stevens H.A. & Heber, R. (eds.), *Mental Retardation,* The University of Chicago Press, 1964.
Thorndike, R.L. *The Concepts of Over- and Underachievement.* New York: Teachers College, Columbia University, 1963.
Tizard, J. Longitudinal and follow-up studies. In Clarke, A.M. & Clarke, D.B. (eds.) *Mental Deficiency. The Changing Outlook,* (2nd ed.), New York: The free press, 1965.
Tizard, J. & O'Connor, N. The employability of high-grade mental defectives, I. *Amer. J. Ment. Defic., 54:* 563–576, 1950(*a*).
– The employability of high-grade mental defectives, II. *Amer. J. Ment. Defic., 55:* 144–157, 1950 (*b*).
Trankell, A. *Den laterala dominansen. En studie i beteendets etiologi.* Pedagogiska institutionen, Stockholm, 1960.
Wagner, E.E. & Hawver, D.A. Correlations between psychological tests and sheltered workshop performance for severely retarded adults. *Amer. J. Ment. Defic., 69:* 685–691, 1965.
Vernon, P.E. Educational testing and test-form factors. *Research Bull. 58–3. Princeton: Educational Testing Service,* 1958.
Wessman, L. & Gustin, T. *Efterundersökning av särskolelever födda åren 1935–1939, del I och II.* Published by Kungl. Skolöverstyrelsen, Stockholm (mimeographed), 1964.
Windle, C.D. Prognosis of mental subnormals. *Monogr. Suppl. Amer. J. Ment. Def. 66:* 1–180, 1962.
Wolfensberger, W. Vocational preparation and occupation. In Baumeister, A.A. (ed.) *Mental Retardation,* University of London Press Ltd., 1967.
Österling, O. *The Efficacy of Special Education,* Svenska Bokförlaget/Norstedts, Uppsala, 1967.

In this book studies are reported concerning the possibilities of predicting performance at work and adjustment to work in mentally retarded adults by psychological tests and observational methods.

After reviewing earlier efforts of this kind, the author describes the prediction variables used, viz., measurements of psychomotor, spatial-inductive and numerical abilities and also observations of behaviour in general work situations. The criterion is defined as a primarily quantitatively determined capacity to carry out approved tasks in a sheltered workshop. During the first stage of the investigations the author procured three subjective ratings of this capacity which proved to be highly reliable. Later, eight tasks were selected which provided objective measures of the criterion variable. Finally, the author developed a number of composite criterion measures which are based on the measurements and systematic observations made in connection with the eight tasks.

The predictive power of the abilities and of the behavioural observations was determined by repeated stepwise multiple-regression analyses. The results show convincingly that combinations of the above-mentioned variables can function as valid predictors, regardless of the criterion measures used—subjective ratings, objective measures of performance on individual tasks or composite measures of performance and adjustment.

It is of special interest that the author is able to demonstrate a certain dimensionality in the work performance of the mentally retarded and that he finds an apparent relationship between the nature of the tasks, as revealed by factor analyses. and their predictability. Thus, it seems to be possible to make more accurate predictions of performance on long-cycle and complex tasks, which also require the handling of tools, than of corresponding performance on short-cycle and simple tasks.

ISBN 91-24-25148-8
Best.-nr 24-25148-8
(24-25148-8) B